問題解決の視点からみた
現在・過去・未来（Ⅶ）

岡　山　　誠　司

目　次

3　情報の量とその計測法

情報とか，情報化というコトバは今日では聞きなれてしまって“あいさつ”なみになってしまった．しかしコンピュータや各種の通信機器の急速な進展がみられるようになったのは通信（情報）理論が確立されたことがその背景にある．それはあらゆる情報表現や表記をすべて2元符号語に変換して伝送することを前提とした理論大系であった．従って情報の量は，ビット単位で表わすことで統一されてしまった．

ここでは，通信やコンピュータで使われるビット単位の他に，日常的な，社会的な意味で受取られ，理解されている，そうした情報にまで視野を広げて“量”の概念をサーベイし，そうしたなかで，いわゆる“ビット”単位はどのように位置づけられるのかを探っていく．

このような立場から，情報の計測法をレベルの低いほうから以下，順次とりあげていく．

（A）　容器法

情報を入れる「入れもの」の大きさではかる法である．

例えば，新聞紙という情報を運ぶ入れものがある．この入れものは，平日の新聞よりは日曜日の新聞が，また朝刊や夕刊だけよりは，朝・夕刊のセットで購読するほうが入手できる情報量が多い．タブロイド版よりふつう版のほうが情報量が多い．

ラジオも音声―電波―音声という形式で情報を耳へ伝える．夜の 11:00 で放送終了する放送局よりも朝の 2:00 まで放送する民間局が送ってくる情報量は多い．このような認識のしかたで情報の量を理解するばあいである．

現実には，こうしたはかりかたが結構採用されている．

（1） 郵便

はがき：1枚分40円．字を細かくしていくら多量の文字を書いても料金は割増にはならない．しかし1枚分を“私製はがき”だからといって勝手に大きなサイズにはできない．

許される絵はがきのサイズは長辺14～15 cm，短辺9～10.7 cm と制限されている．

定形郵便物なら長辺14～23.5 cm，短辺9～12 cm，厚さ1 cm以下，重量50 g以下など．

てがみ：重さ（g数）で料金がきまる．定形郵便物なら25 gまで60円，50 gまで70円など．

電子（コンピュータ）郵便：郵政省製の封筒と通信文用紙が1枚ついて20円，これに郵送料60円を加えて80円（注：'89・3・31以前の料金体系）

（2） 電話

これも“秒数”で料金がきまっており，早口で多くの言葉をしゃべったからといって料金は増額するといったシステムではない．

このほか，旅行先での案内ガイド料金は，例えば1日当り10ドル，というきめかたがふつうである．

しかし，こうした情報をのせる“空間”や“時間”の“入れものや大きさ”をもって，そこに盛りこまれた情報の量とするのは荒っぽいきめかたではある．

スクラップ（切張り）用紙にはった1枚分の情報といっても，それが，ふつうの新聞紙からの切抜きのばあいと縮刷版のコピーのばあいとで，またマイクロフィルムが収納されているばあいとでは，情報量は明らかに後者ほどふえているはず．逆に点字や大活字本の切抜きでは減ってくる．

そこで，**情報面密度**で，まず比較することになる．

新聞1ページ分（面積 A）の情報量を I とすれば，縮刷版ないしマイクロフィルムに縮小したとき，その倍率を β とすると $\beta \equiv l/L$

ここで L: オリジナルの辺長，l: 縮刷の辺長である．縮刷の面積を a とすれば，単位面積当りの情報量，つまり情報面積密度 σ は新聞（オリジナル）で $\sigma_A = I/A$，縮刷で $\sigma_B = I/a$，ここで $a/A = (l/L)^2 = \beta^2$，従って，情報面密度の比，つまり情報面密度倍率 σ_B/σ_A は，大きさ倍率 β の2乗の逆数となる．

情報時間密度についても同様に考えられる．

標準速度で1本分を時間 T で吹きこんだ情報量を I とすれば，これを圧縮したとき，その倍率を γ とすると $\gamma \equiv t/T$

ここで t は圧縮したテープの収録時間である．

単位時間当りの情報量，つまり，情報時間密度 ρ は，標準テープで $\rho_A = I/T$，圧縮テープで $\rho_B = I/t$,

従って，情報時間密度の比，つまり情報時間密度倍率 ρ_B/ρ_A は，時間倍率 γ の逆数となる．

このように情報密度のような基準化を行ってみると，単なる"入れもの"で情報の量をはかることは混乱のもとになることがわかる．そこで，つぎは，情報の量を"文字数"ではかることになる．なおここでは，文章またはそれに等価換算できる話しコトバに，情報を限定しておく．

（B）　文字数法

情報を構成している文字列に含まれる文字数の大きさ（文字列の長さ）ではかる法である．

例えば，1月当り10冊の雑誌を読むばあい，そこに含まれている情報量を，

10（冊/月）×100（ページ/冊)1000（語/ページ）× 8（字/語）

$=8\times10^6$（字/月）

と計算して，月あたり800万字の情報量がこれらの雑誌により提供されたとみる．年間には，この12倍，約1億字とみつもる．

経企庁の予測（1985）によると，テレビ，ラジオ，新聞，雑誌，書籍，郵便，電報，電話，テレビ電話，コンピュータなどから与えられる，国民1人当りの情報量は，1年間で，前出の1億字の約40倍（A 5，200ページの雑誌2万冊分）になる（なおこれには人と人との直接対話の情報は含まれていない）と新聞に発表していたが，このはかりかたの例である．

とにかく，このように文字数で情報の量の目安とする考え方も，広く現実に採用されている．

（1） 電報

字数によって料金がきまる．

（2） レタリング

1文字ごとで料金がきまる．

このように，文字数で情報の量をきめるとき，文字のコ数の単なる多さだけにとどまらず別の意味が生じてくる．

文章で，文字数が多くなると一般に，“区別”がより細かくできるようになる．

情報であることの基本条件は，他とそれとが“区別”できることにある．

例えば，

①「ネコ」がいる……という文章だけではそこにネコがいるというだけ，つまり，そこにいるのは，ライオンでもネズミでもないということを指示している．

② 隣りの家の「ネコ」がいる……となれば，向いの家のネコでも後ろの家のネコでもないことが明らかになる．ネコの飼主がはっきりする．

③ 隣りの家の三毛の「ネコ」がいる……となれば，ネコは白でも白黒ブチでも

ない，とネコの外観が区別されてくる．

④ 隣りの家の三毛のおやネコがいる……となるとネコは子ネコではないと指定される．

このように区別がよりくわしくなっていくほど，“情報”は増えたことになる．そのためにはふつう文字数をそれだけ多くしなければならない．

では，情報の量を文字数で表わせば，それで問題がないか，といえば，そう簡単ではない．

日本文で，漢字かなまじり文をすべてかな書きにすると，文字数はふえる．

つまり，同じ意味内容（情報量）が，異なる文字体系では異なる文字数になってしまう．もちろん，日本文を外国語へ翻訳する際にも同じことがおこる．

例えば，

ポラロイド・ランドフィルム　Type 108 のパッケージの中に取扱説明書がはいっている．取扱い方，保証についての説明，取扱い上の注意事項の記載が，世界 12 カ国語でなされている．

表 1　ポラロイド・ランド・フィルム(Type 108)保証と注意の文字数

国　語	文字数
Italiano	940
Dansk	858
Português	857
Deutsch	840
Nederlands	815
Espãnol	814
Français	804
English	774
Svenska	739
日本語	380
中国語	199

（注：アラビア語は除く）

いま，これらの記載事項のうち，保証と注意の部分だけまとめて文字数を調べてみた．

文字数の多い国語からならべてみる（表 1)．

最も短いのは，いうまでもなく中国語，これを規準にして日本語（漢字かなまじり文）が約 2 倍の長さになる．欧米語ではスウェーデン語が最短で，これを規準にしてイタリー語はその 3 割増しになっている．この違いは 1/e 法則に照らしても無視できない．

図　1

外国語教室

	英語	ドイツ語	日本語		英語	ドイツ語	日本語
1	BAT（バット）	die Fledermaus（フレダーマウス）	こうもり	24	M. DUCK（マンダリンダック）	die Brautente（ブローテンテ）	おしどり
2	BEAR（ベア）	der Bär（ベア）	くま	25	MONKEY（モンキー）	der Affe（アッフエ）	さる
3	CAMEL（キャメル）	das Kamel（カメル）	らくだ	26	OX（オックス）	der Stier（ステイーア）	おうし
4	CAT（キャット）	die Katze（カッツエ）	ねこ	27	PANDA（パンダ）	der Panda（パンダ）	ぱんだ
5	COCK（コック）	der Hahn（ハーン）	おんどり	28	PARROT（パロット）	der Papagei（パパガイ）	おうむ
6	COW（カウ）	die Kuh（クー）	めうし	29	PEAFOWL（ピーフオール）	der Pfauhahn（プフアウハーン）	くじゃく
7	DEER（デイア）	der Hirsct（ヒアシユ）	しか	30	PENGUIN（ペンギン）	der Pinguin（ピングイーン）	ペンギン
8	DOG（ドック）	der Hund（フントウ）	いぬ	31	PELICAN（ペリカン）	der Pelikan（ペリカーン）	ぺりかん
9	DUCK（ダック）	die Ente（エンテ）	あひる	32	PIGEON（ピジョン）	die Taube（タウベ）	はと
10	EAGLE（イーグル）	der Adler（アドウラー）	わし	33	POLAR・BEAR（ポーラーヘア）	der Nördpol Bär（ネーデポルベア）	ほっきょくぐま
11	ELEPHANT（エレファント）	der Elefant（エレファント）	ぞう	34	PORCUPINE（ポーキユパイン）	das Stachelschwein（シユタツハルシユワイン）	やまあらし
12	FOX（フオックス）	der Fuchs（フツクス）	きつね	35	RAT（ラット）	dir Ratte（ラテ）	ねずみ
13	FURSEAL（フアーシール）	der Seebär（ゼーベア）	おっとせい	36	RABBIT（ラビット）	der Hase（ハーゼ）	うさぎ
14	GOAT（ゴウト）	dir Ziege（チーゲ）	やぎ	37	RACCOON（ラクーン）	der Waschbär（ヴアシユベア）	あらいぐま
15	HAWK（ホーク）	der Falke（フアルケ）	たか	38	RHINOCEROS（ライノサラス）	der Nashorn（ナシヨールン）	さい
16	HEN（ヘン）	dir Henne（ヘネ）	めんどり	39	SHEEP（シープ）	das Schaf（シヤーフ）	ひつじ
17	HIPPO（ヒポウ）	das Nilpferd（ニールプフエアト）	かば	40	SPFRROW（スパロー）	der Spatz（スパッツ）	すずめ
18	HORN-OWL（ホーンアウル）	die Ohreule（オーロイレ）	みみずく	41	SQUIRREL（スクイラル）	das Eichhörnchen（アイヒヘルンヒエン）	りす
19	HORSE（ホース）	das Pferd（プフエアト）	うま	42	TAPIR（テイパー）	der Tapir（タピール）	ばく
20	LEOPARD（レパード）	der Leopard（レオパード）	ひょう	43	TIGER（タイガー）	der Tiger（テイガー）	とら
21	LION（ライオン）	der Löwe（レーヴエ）	らいおん	44	TORTOISE（トータス）	dir Schildkröte（シルトクレータ）	かめ
22	LYNX（リンクス）	die Wildkatze（ヴイルトカッツエ）	おおやまねこ	45	WILDBOAR（ワイルドボー）	das Waldschwein（ヴイルトシユバイン）	いのしし
23	MACAW（マコー）	der Sittich（ジチイッヒ）	いんこ	46	WOLF（ウルフ）	der Wolf（ヴオルフ）	おおかみ

（ギンピス　たべっこどうぶつ）

中国語の漢字と，日本語の漢字かなまじり文でも長さに差がつくのは当然であるが，欧語どうしでも結構な違いがあることがわかる．

別の例：「ビスケット」のパッケージ外側に46種の動物が英語，独語，日本語（かな）で並記してあるのが眼にとまった（図 1)．1単語当りの平均字数を求めると英語で5.48字．独語で6.35字，日本語で3.15字となる．

これより英語は日本語の1.74倍，これは前出のフィルムの例のばあい774/380＝2.04にほぼ近い．

また独語は英語の1.16倍，これもフィルムの840/774＝1.09にほぼ近い．

単純な比較であるが，われわれの言語系の相違は，こうした日常的に眼につくサンプルからイメージしてることを考えるなら気軽に数値的なチェックをしてみることの必要性を痛感する．

（C）　文字列種数法

ここで，文字列のもつ，もう1つの別の側面に眼を向けてみよう．

さきに，同じ内容が異なる文字数になると気づいたが，逆に，同じ文字数でどれだけ異なる内容が表わせるかを考えてみる．ふつう内容を異にするためには，文字列が異ならなければならない．これは必要条件である．しかし，意味ある文章の各文字を入れかえてつくれる異なる文字列は，必ずしも異なる別の意味内容をもつようになるとは限らない．

大部分は意味なき単なる文字列に過ぎなくなる．

しかし，第1近似として，文字を入れかえてつくれる異なる文字列の種類数（種数）が異なる内容の文字列をつくる可能性があるとして，情報の量の指標とすることが考えられる．つまり，情報を構成している文字列のもつ情報量を，各文字を入れかえてつくれる異なる文字列の種類ではかる法である．

文字数 n コの文字列を考える．

①文字数nが多くなると，一般に異なる文字列の種類は大きくなる．これより，この情報量は文字数法ではかる情報量とその方向が一致することがわかる．そのほかこの法では，

②文字（アルファベット）の種数（**字種数**）M が大きくなっても，文字列の種数（**列種数**）は大きくなる特長をもつ．

そこで，字種数 M で，長さ n の文字列がもつ情報の量は，列種数 M^n で与えられると定義する．

欧語のアルファベットについては，英語では A から Z まで $M=26$ だが，独語では Ü, Ä, Ö およびエスツェット β を別字とすると $M=30$ となるし，仏語では ç；é；à, è, ù；â, ê, î, ô, û を別字とすると $M=36$ となる．

スペイン語は Ch, Ll, Ñ, rr を追加して $M=30$，ギリシャ語では $M=24$，露語では $M=33$，イタリヤ語では，外国語を示すための5種を追加して $M=26$，このほか，アラビア語で $M=29$，ポリネシア語で $M=16$ などがある．

字種数 M が大きいとき情報量が大きくなるという考え方は，タイプライターで，かなうち専用の機種にくらべて漢字打ちのできる和文タイプの価格が高くなるのもなっとくできる．

情報量は，文字長nと字種数Mとで構成されると分解できるが，ここで，

①文字長 n を大にするときは，「こまかい限定ができる」とされ，

②字種数 M を大にするときは「表現のバライティが増やせる」ともいわれている．

前者の例としては，「ネコ」を使って説明した（前出）が，後者の例としては，

わたしはおどろきました

私は驚きました

私はギョッ!! とした

などを比較してみればうなづけよう．

日本語は，かな，カナあわせて約100種，これに一般に使わる漢字2000種を加わえてわかるように欧語にくらべ字種はケタ違いに大きい．それだけ微妙な表現ができるわけである．

列種数による情報量のきめかたは，文字の長さ n だけで決めるばあいを含み，且つそれの拡張であると位置づけられる．

さて，この情報量 M^n は，M や n が大きくなるとその値は急増する．そこで対数をとり

$$\log I = n \log M$$

ここで　　$\log I$

を，**多様さからみた情報量**とよぶ．

これを使うと，文字１コ当りの情報量（**平均情報量**，または**エントロピー**ともよぶ）H は

$$H \equiv \log M$$

となる．

ここで，字種数の変更が文字列の長さにどのような影響を与えるかをみておく．

字種数 M で文字長 n の文字列に対して，

字種数 M' で文字長 n' の文字列が，同じ情報量をもつとすると

$$n \log M = n' \log M'$$

これより，長さの比 n'/n は

$$n'/n = \log M/\log M'$$

となる．

いま $M'=2$（２元符号系）とすると，英語では $M=26$ だから

$$n'/n = \log 26/\log 2 \fallingdotseq 4.70$$

つまり英語文を２元符号に直すと，約５倍の長さに伸びることがわかる．

なお，情報量を求めるときの対数は，底を10にとる，e にとる，2にとる，などいろいろ考えられるが，2元符号を基準にとることを考えるなら底を2にすると $n'\log_2 2=m$ で，情報量は，2元符号化したときの長さそのものとなる．底を2つにとった対数を lg で表わし，単位を**ビット**(bit)とよぶ．

クルマのナンバープレートの情報量

例えば，日本式の「さ- 2365」がもつ情報量と，カリフォルニアの「VTQ-23」がもつ情報量を比較してみるといずれも5文字列（5ケタ）で同じだが，かなと，アルファベットでは字種数が異なることに注意して，日本式では，1字目のかな44種，残り4ケタの数字，各ケタ10種だから情報量は 44×10^4，これより $\lg 440000=18.75$ bit，つまり高々19ビット，これに対して米国式では1字目から53字目まで，各26種だから $\lg(26^3\times10^2)=20.75$ bit，約21ビットに達する．これより後者の表示形式のほうが，含有する情報量が約2ビットつまり約4倍多いといえる．

チャオ　『言語学入門』岩波書店　1980

(D)　いいあて法

情報を構成している文字列──与えられた特定の文字列──に注目する．その文字列を“特定”あるのに必要な“コスト（費用)”の程度ではかる法である．

コストが余計かかる文字列ほど，情報量が大きいことになる．

ここで“特定する”とは，自分がその文字列を知ることができ，また他人にも伝えられることを意味する．また，“コスト”とは，文字列を知るにいたるまでの困難さや努力の程度を計量的に表わしたものを指す．」

基準モデルとしての「数字札あてゲーム」

数字札1から10までを使う．十分にシャッフルしたあと1枚だけ抜き出し，その数字を被験者にいいあてさせる．ただし，いいあてる手続き(ルール)としては，つぎのような「質問回答法」による．

被験者の「それは……ですか」という質問に対し，実験者は「Yes」かまたは「No」だけで回答する．

このときコストは，被験者がいいあてるまでに続けた質問の“回数”とする．

さて，文字列の最初の文字からいいあてていくが，第1文字をいいあてるための質問回数を q_1，一般に i 番目の文字について質問回数を q_i とするとすべての文字列をいいあてるための総質問回数 Q は

$$Q=\sum_{i=1}^{n}q_i$$

となる．これを情報量とする．

一般化すると，ある文字列があるとして，これを全く知らない人が，知っている人から1字づつ知らされる状況を設定すれば，上記のような質問回答法によって1字づつ知っていくという過程をイメージしていることになる．この情報量は従って一連の n コの文字列の“知りにくさ”の程度を表わしているともみらる．

情報量を決めるための条件としてその文字列に含まれる字種はあらかじめ知らされているとする．このとき，情報量の性質，つまり Q は

① Mが大きくなれば，また

② n が大きくなれば，いずれも大きくなるであろう．

この情報量は，前出した文字列種数法で定義した情報量とその傾向は一致する．

そこで，もう少し，くわしく M と n そして質問回答のしかたと，質

問回数 q_i の関係を調べてみる．

① $M=1$ （数字札1だけしかないばあい）これをあらかじめ知らされているのだから，質問するまでもなく，「その札は1である！」といいあてられる．つまり質問回数 $q_1=0$，これを n 回つづけても $Q=0$

平均情報量 $\bar{q}$ も

$$\bar{q}\equiv\frac{Q}{n}=0$$

となる．

② $M=2$ （数字札は1と2だけであるばあい）これを知らされているとして，質問は，最も単純なばあいは「1ですかと質問する．yes と回答があれば，札の数字はもちろん 1，no と回答されるなら，札の数字は2と特定できる．この質問手順と回答の関係をフローチャーと（図2）に示す．

こうして，札の数字が1であれ2であれ質問回数 $q_1=1$ となる．これを n 回つづければ $Q=n$，従って $\bar{q}=1$ となる．

③ $M=3$ （数字札は 1, 2, 3 の3種であるばあい）これを知らされているとして，最も単純な質問手順とみられるのは（図3）に示す．

今度は，札が1か，2または3かでは質問回数が異なる．札が1なら質問は1回，2または3なら質問は2回となる．

文字列の長さ n コとして，このなかに札1が n_1 コ，札2が n_2 コ，札3が n_3 コ含まれていたとすると，いいあてるに要する総質問回数 Q は

$$Q=\sum_{i=1}^{3}n_iq_i$$

ここで $q_1=1, q_2=q_3=2$ である．

また $\bar{q}=\dfrac{Q}{n}$ ここで $n=\sum_{i=1}^{3}n_i$

これより Q または $\bar{q}$ は，n が同じであっても n_1, n_2, n_3 の分配の様子

図 2　質問のフローチャート (M=2)

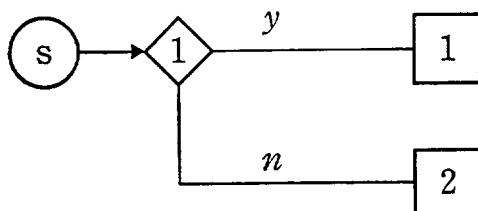
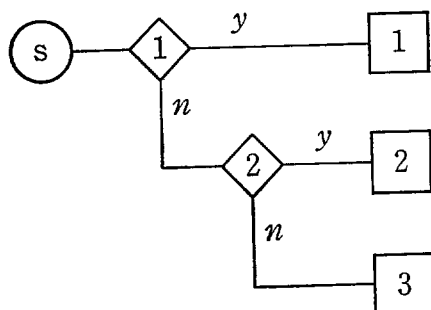



図 3　質問のフローチャート (M=3)

で変わってくる.

いま n_1, n_2, n_3 が“等分配”であったとすると

$$n_1/n = n_2/n = n_3/n = 1/3$$

だから

$$\bar{q} = \frac{1}{3}\times 1 + \frac{1}{3}\times 2 + \frac{1}{3}\times 2 = \frac{3}{5} \fallingdotseq 1.67$$

となる.

なお，質問の手順として，最初に 1 ですかとするかわりに 2 ですかでも 3 ですかでも数字が変わるだけで，フローチャートの様相は変わらないから，これ以外の異なる質問の手順は考えなくてよい.

④　M=4 (数字札は 1, 2, 3, 4 の 4 種であるばあい) これを知らされているとして，最も単純な質問手順とみられるのは(図 4)，これは M=3

図 4　質問のフローチャート（$M=4$）

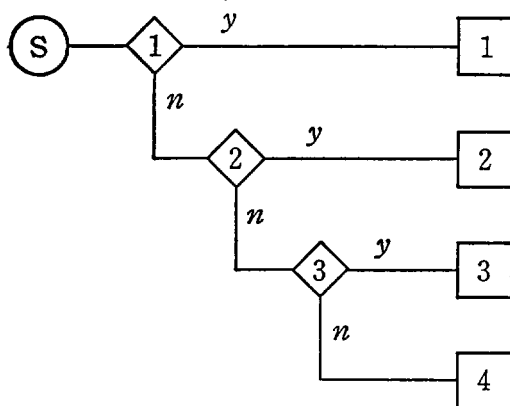
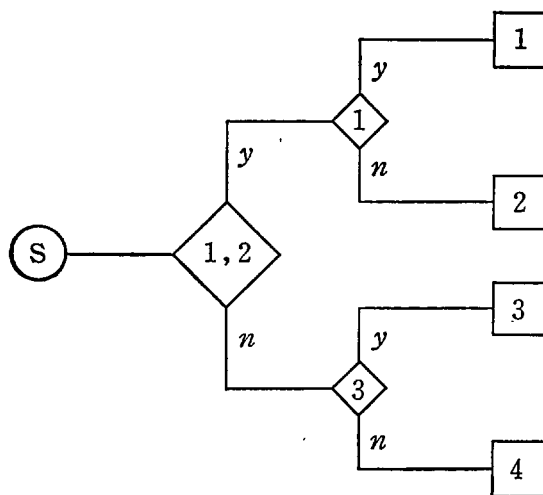



図 5　質問のフローチャート（$M=5$）

の質問手順の単なる拡張である．そこで

文字列の長さ n コとしたとき

$$Q=\sum_{i=1}^{4} n_i q_i$$

ここで $q_1=1, q_2=2, q_3=q_4=3$ である.

また $\bar{q}=\dfrac{Q}{n}$, ここで $n=\sum_{i=1}^{4} n_i$

いま, n_1, n_2, n_3, n_4 が“等分配”であったとすると

$$n_1/n=n_2/n=n_3/n=n_4/n=1/4$$

だから

$$\bar{q}=\frac{1}{4}\times 1+\frac{1}{4}\times 2+\frac{1}{4}\times 3+\frac{1}{4}\times 3=\frac{9}{4}=2.25$$

となる.

今度は, 別の質問の手順も考えられる. (図5)

最初の質問は, 今度は「1か2ですか」となっている. このかわりに「2に等しいか, それ以下ですか」と質問してもよい.

文字列の長さを n コとしたとき

$$Q=\sum_{i=1}^{4} n_i q_i$$

ここで $q_1=q_2=q_3=q_4=2$ であるから

$$Q=2n$$

となり, 等分配であろうとなかろうと関係なく n だけできまる.

また $\bar{q}=\dfrac{Q}{n}=2.00$

となり, これは, さきの質問手順のばあいにくらべて“小さく”なる.

質問の手順は以上の2種しかない.

従って, 後者の質問手順によれば q は最小になるといえる. 質問回数はコストであるから以後はコストを“最小”にするばあいの質問手順を“基準”として採用することに約束する.

最小値を基準に採用する理由は, ムダな質問を許すと質問回数は“とめどもなく”増えていくことが予想されるからである.

賢い人は，情報を“最小コスト”で入手するが，ボンヤリした人は，同じ情報をコスト高で入手することになる，という事情を思い合わせればなっとくがいくであろう．

この後者の質問の手順（**等分法**）は，任意の字種数 M には適用できないが，$M=2^r$ $(r=1, 2, \cdots)$ のばあいには拡張して使える．

このときは，常に $\bar{q}$ は最小値となり

$$\bar{q}=\frac{Q}{n}=\frac{1}{n}\sum_{i=1}^{M}n_i q_i=q$$

ここで

$$q=r=\lg M$$

である．

これに対して，前者（図４）の質問の手順（**順序法**）は，任意の字種数 M に拡張して適用できる（図６）．

文字列の長さ n コとしたとき

$$Q=\sum_{i=1}^{M}n_i q_i$$

ここで $q_1=1, q_2=2, q_3=3, \cdots\cdots q_{M-1}=q_M=M-1$

である．

ここで等分配ならば

$$n_1=n_2=\cdots\cdots=n_M=n/M$$

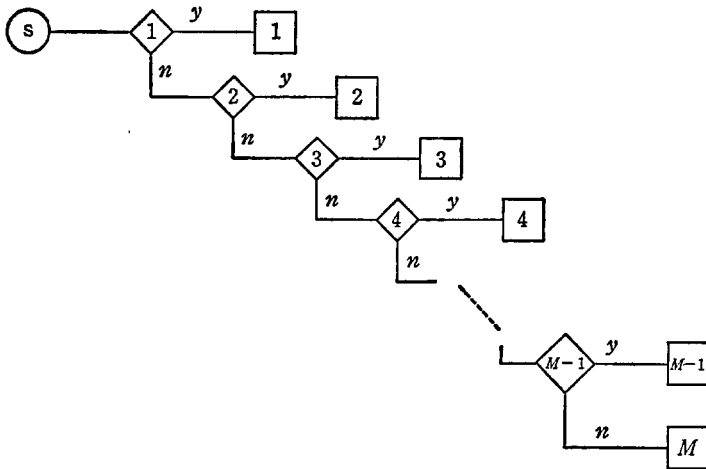
だから

$$Q=\frac{n}{M}\left[\frac{M}{2}(M+1)-1\right]$$

$$=n\left[\frac{M+1}{2}-\frac{1}{M}\right]$$

そして

$$\bar{q}=\frac{Q}{n}$$

図 6　質問のフローチャート（M）

で求められる．

このように $\bar{q}$ は，等分配のばあいは，順序法で質問すると，等分法で質問するばあいよりその値は大きくなるが，等分配でないばあいは，出やすい文字がわかっているなら，逆に等分法で質問するばあいより“小さく”なることは十分に考えられる（これについては後述する）．

さて，等分配であるが，字種数 M が $M \neq 2^r (r=1, 2, \cdots\cdots)$ のばあいの検討をすませておこう．

$M=3$ では，等分法は採用できない．順序法では（図 3）$\bar{q}=5/3=1.67$ で，これより小さい値にはならない．

$M=5$ でも等分法は採用できない．しかし，等分法に近い分割を採用することで順序法とは異なる質問の手順が考えられる．

それをもうら的に取りあげ，等分配のばあいの $\bar{q}$ を求めてみる（図 7,

図 7　質問のフローチャート（$M=5$）

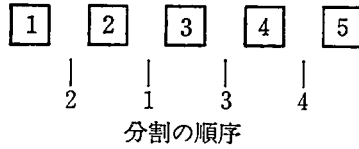
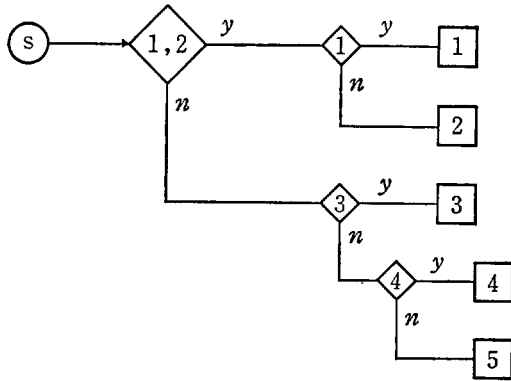
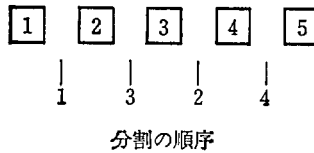
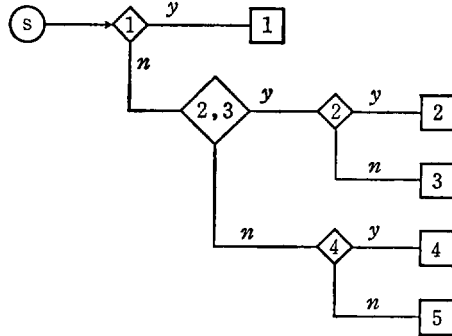
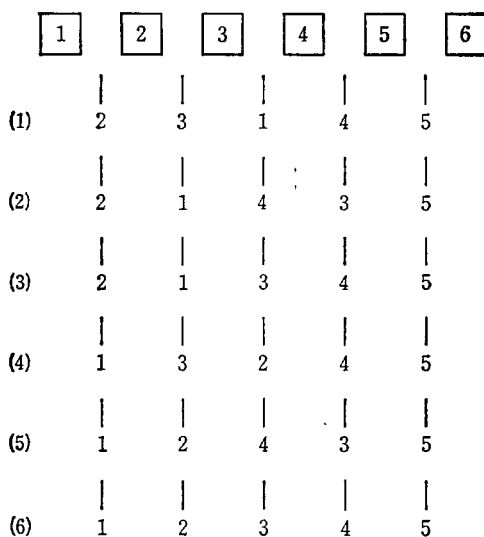



図 8　質問のフローチャート（$M=5$）

図 9　質問のための分割の順序（$M=2$）

	1		2		3		4		5		6
(1)		2		3		1		4		5	
(2)		2		1		4		3		5	
(3)		2		1		3		4		5	
(4)		1		3		2		4		5	
(5)		1		2		4		3		5	
(6)		1		2		3		4		5	

図 8）.

これより，等分配のばあいは，等分法に最も近い分割にもとずく質問の手順（**等準法**）を採用するとき $\bar{q}$ が最小になることがわかる．

$M=6$ のばあいについても，異なる質問の手順のすべてについてとりあげ，等分配のばあいの $\bar{q}$ を求めた（図 9）．ただしフローチャートは省略する．

$M=7$ のばあいについても，同じように求めてみた（図 10）．

以上を表にまとめた（表 2）．

これより，各 M について $\bar{q}$ の最小値を取出すと，この数値は，ある規則的な数列になっていることがわかる．

この数列は，分数で表わしたとき，分母は M，分子は，$M=2^k+1$ から 2^{k+1} までの M については，M が 1 ますごとに，分子の差は常に $k+2$ になっている．

表 2　字種数と平均情報量

字種数 M	$\bar{q}$ 等分法 ($\lg M$)	$\bar{q}$ 等分法，順序法以外の法	$\bar{q}$ 順序法
1	0.00		
2	1.00		
3	1.58		5/3=1.67
4	2.00		9/4=2.25
5	2.32	(1) 12/5=2.40 (2) 13/5=2.60	(3) 14/5=2.80
6	2.58	(1), (2) 16/6=2.67 (3) 17/6=2.83 (4) 18/6=3.00 (5) 19/6=3.17	(6) 20/6=3.33
7	2.81	(1) 20/7=2.86 (2), (3) 21/7=3.00 (4) 22/7=3.14 (5), (6), (7) 23/7=3.29 (8) 24/7=3.43 (9) 25/7=3.57 (10) 26/7=3.71	(11) 27/7=3.86
8	3.00		35/8=4.72
⋮	⋮	⋮	⋮

ここで，1コの文字を完全にいいあてるまでに必要な“最大質問回数”を $\bar{q}_{mm}$ とすると（質問の手順は等分法か等準法を採用するとして）

$M=2^k+1$ から 2^{k+1} までの M については

$$\bar{q}_{mm}=k+1$$

と推定できる．これを

図 10　質問のための分割の順序（$M=7$）

	1		2		3		4		5		6		7
(1)		2		3		1		5		4		6	
(2)		2		3		1		4		5		6	
(3)		2		1		4		3		5		6	
(4)		2		1		3		5		4		6	
(5)		2		1		3		4		5		6	
(6)		1		3		4		2		5		6	
(7)		1		3		2		5		4		6	
(8)		1		3		2		4		5		6	
(9)		1		2		4		3		5		6	
(10)		1		2		3		5		4		6	
(11)		1		2		3		4		5		6	

$$\bar{q}_{mm}=\lfloor \lg M \rfloor$$

と書く．記号 $\lfloor\ \rfloor$ は，$\lg M$ が整数なら，そのままであるが，小数値なら，それより大きい，最も近い整数（切り上げ整数）を意味させる．

最大限この値 $\bar{q}_{mm}$ を見込んでおけば，この質問回数以内に，必ず文字はいいあてられることになる．

さきに情報量を（C）で $n \log M$ で定義したが，ここで底を 2 にとると，1 文字当りの情報量は $\lg M$ となるから，この値は $\bar{q}$ の最小値にほとんど一致する．ただし等分配のばあいについてであるが．

こうして（C）での情報量と，ここ（D）での情報量は，ほぼ一致する

ということで両者での情報量の定義は連絡する．

等分配（等確率発生）でないばあい

このばあいの $\bar{q}$ の最下値 $\bar{q}(\min)$ や，最大質問回数 $\bar{q}_{mm}$ について検討する．

文字列の長さ n コとしたとき，各文字の発生確率が与えられたばあいについて，まず，異なる分割形式のすべてについて $\bar{q}$ を求めて，そのなかで最小となる $\bar{q}(\min)$ を探し出してみる．

いま，数字 i の札の発生確率を $p_i(i=1, 2, \cdots\cdots)$ とする．

$M=3$ のばあい

$$p_1=3/6, p_2=2/6, p_3=1/6$$

で与えられたとする．

これを札1から数字の順に順序法で質問するなら

$$\bar{q}=\frac{3}{6}\times 1+\frac{2}{6}\times 2+\frac{1}{6}\times 2=\frac{3}{2}=1.50$$

なお，等分配では $\bar{q}=\dfrac{5}{3}\doteqdot 1.67$ であったから，明らかに今度は“小さく”なる．

この結果は，常識的にもうなづける．

絶えず発生している事件についての報道は，これを聞かされても感激度ないし興奮度は少ない．つまり，情報がもたらす情報量は減少するのである．

これより，$M=3$ では，特定の数字の札——例えば1——が他の数字2や3の札よりも出現確率が高いことを知っておれば，第1回目に「それは1であるか」と質問することで $\bar{q}$ は減らせる．

$M=4$ のばあい

$$p_1=4/10, p_2=3/10, p_3=2/10, p_4=1/10$$

で与えられたとする．

これを，札1から数字の順に順序法で質問するなら

$$\bar{q}=\frac{4}{10}\times1+\frac{3}{10}\times2+\frac{2}{10}\times3+\frac{1}{10}\times3=\frac{19}{10}=1.90$$

なお，等分配では $\bar{q}=2.0$ であったから今度も明らかに“小さく”なる.

$M=5$ のばあい

$$p_1=5/15,\ p_2=4/15,\ p_3=3/15,\ p_4=2/15,\ p_5=1/15$$

で与えられたとする.

これを，札1から数字の順に順序法で質問するなら

$$\bar{q}=\frac{5}{15}\times1+\frac{4}{15}\times2+\frac{3}{15}\times3+\frac{2}{15}\times4+\frac{1}{15}\times4=\frac{34}{15}\doteqdot2.27$$

となるが，このばあいが $\bar{q}(\min)$ とはならない.

この他の分割にもとずく質問の手順のすべてについて検討してみなければならない（図 7, 8).

このばあいは等準法で q (min) $=2.20$ がえられることがわかる.

$M=6$ のばあい

$$p_1=6/21,\ p_2=5/21,\ p_3=4/21,\ p_4=3/21,\ p_5=2/21,\ p_6=1/21$$

で与えられたとする.

異なるすべての分割にもとずく質問の手順を検討してみる（図 9).

順序法では $\bar{q}=2.62$ であるのに対して $\bar{q}(\min)=2.43$ がえられる分割法があることがわかる.

$M=7$ のばあい

$$p_1 7/28,\ p_2=6/28,\ p_3=5/28,\ p_4=4/28,\ p_5=3/28,\ p_6=2/28,\ p_7=1/28$$

で与えられたとする.

異なるすべての分割にもとずく質問の手順を検討してみる（図 10).

順序法では $\bar{q}=2.96$ であるのに対して今度も $\bar{q}(\min)=2.64$ がえられる分割法のあることがわかる.

さて，この程度の検討からは，一般に文字の発生確率が与えられたとしても，どのような分割形式を採用すれば $\bar{q}(\min)$ がえられるかという方針は発見できない．

ただ，傾向としては，p_i が大であれば q_i が小となるような分割法形式にもとずき質問の手順を決定するのが $\bar{q}$ を小さくするための方策となる，と主張できる．

この方策を，つぎに追跡してみる．

$\bar{q}(\min)$ をうるための探索

これまでは，与えられた種数 M についてその分割形式のすべてをもうら的に取上げて，それにもとずく質問の手順を適用して $\bar{q}$ を求め，そのなかから $\bar{q}(\min)$ を選出していた．

ここでは，逆に与えられた $p_i(i=1,2,\cdots)$ を使ってあるルールを適用して q_i を決定する．そのような質問の手順であるかどうかは，この方法で求められる $\bar{q}=\sum p_iq_i$ が $\bar{q}(\min)$ にどの程度近接するかを吟味した上で，あらためて考えることにする．

このように，p_i から導かれた q_i を使用するとき，これまでの q_i と区別するために $\hat{q}_i$ を使用する．

なお，これ以後で q_i を，$\bar{q}(\min)$ がえられる分割形式とそれにもとずく質問の手順で行われたときの値とする．

（i） $\hat{q}_i=\lfloor 1/p_i\rfloor$ を採用するばあい

たしかに，この関係は p_i が大になれば $\hat{q}_i$ は小さくなるので，前出の方策に見合うルールと認められる．

これを採用したばあいの $\hat{q}_i$ と q_i とを比較してみる．

$M=4$ のばあい．p_i は前出と同じとする．このとき

$$\hat{q}_1=3,\ \hat{q}_2=4,\ \hat{q}_3=5,\ \hat{q}_4=10.$$

他方 $\bar{q}(\min)$ かえられる等分法を採用すれば $q_1=1, q_2=2, q_3=q_4=3$ な

表 3　発生確率と質問回数（$M=4$）

$\hat{q}_i \equiv \lfloor 1/p_i \rfloor$

i	1	2	3	4
p_i	4/10	3/10	2/10	1/10
$1/p_i$	2.50	3.33	5.00	10.00
$\hat{q}_i$	3	4	5	10
q_i	1	2	3	3

q_i：(2) 順序法

ので，明らかに

$$\hat{q}_i > q_i \quad (\mathrm{i}=1, 2, 3, 4)$$

である（表 3）．

$M=5$ のばあい．p_i は前出と同じとする．このとき

$$\hat{q}_1=3, \hat{q}_2=4, \hat{q}_3=5, \hat{q}_4=8, \hat{q}_5=15$$

やはり

$$\hat{q}_i > q_i \quad (i=1, 2, 3, 4, 5)$$

である（表 4）．

表 4　発生確率と質問回数（$M=5$）

i	1	2	3	4	5
p_i	5/15	4/15	3/15	2/15	1/15
$1/p_i$	3.00	3.75	5.00	7.50	15.0
$\hat{q}_i$	3	4	5	8	15
q_i	2	2	2	3	3

q_i：(1)

$M=6$ の ばあい．p_i は前出と同じとする．このとき

表 5　発生確率と質問回数　（$M=6$）

i	1	2	3	4	5	6
p_i	6/21	5/21	4/21	3/21	2/21	1/21
$1/p_i$	3.50	4.20	5.25	7.00	10.5	21.0
$\hat{q}_i$	4	5	6	7	11	21
q_i	2	2	2	3	4	4

q_i：(3)

$$\hat{q}_1=4, \hat{q}_2=5, \hat{q}_3=6, \hat{q}_4=7, \hat{q}_5=11, \hat{q}_6=21$$

やはり $\hat{q}_i > q_i$ である（表 5）．

$M=7$ のばあい，p_i は前出と同じとする．このとき

$$\hat{q}_1=4, \hat{q}_2=5, \hat{q}_3=6, \hat{q}_4=7, \hat{q}_5=10, \hat{q}_6=14, \hat{q}_7=28$$

やはり $\hat{q}_i > q_i$ である（表 6）．

以上の数値例から，このような $\hat{q}_i$ の決定法は，その結果のもつ性質か

表 6　発生確率と質問回数（$M=7$）

i	1	2	3	4	5	6	7
p_i	7/28	6/28	5/28	4/28	3/28	2/28	1/28
$1/p_i$	4.00	4.67	5.60	7.00	9.33	14.0	28.0
$\hat{q}_i$	4	5	6	7	10	14	28
q_i	2	2	3	3	3	4	4

q_i：(3)

らいって許容できるものではない．

（ii）$\hat{q}_i=\lfloor \lg 1/p_i \rfloor$ を採用するばあい

このような $\hat{q}_i$ の算出方式も，前出の方策を満たす．これを採用したばあいの $\hat{q}_i$ と q_i とを比較してみる（表 7）．

$M=4$ のばあい．p_i は前出と同じとする

このとき，

表 7　発生確率と質問回数（確率は前出）

（$M=4$）

i	1	2	3	4
$\hat{q}_i$	2	2	3	4

（$M=5$）

i	1	2	3	4	5
$\hat{q}_i$	2	2	3	3	4

（$M=6$）

i	1	2	3	4	5	6
$\hat{q}_i$	2	3	3	3	4	5

（$M=7$）

i	1	2	3	4	5	6	7
$\hat{q}_i$	2	3	3	3	4	4	5

$$\hat{q}_1=2,\ \hat{q}_2=2,\ \hat{q}_3=3,\ \hat{q}_4=4$$

となる.

$\hat{q}(\min)$ がえられる順序法の q_i と比較すれば

$$\hat{q}_i>q_i(i=1,2,3,4)$$

である.

$M=5$ のばあい. p_i は前出と同じとする

このとき

$$\hat{q}_1=2,\ \hat{q}_2=2,\ \hat{q}_3=3,\ \hat{q}_4=3,\ \hat{q}_5=4$$

となる.

$\bar{q}(\min)$ がえられるばあいの q_i と比較するなら, 今度は

$$\hat{q}_i\geq q_i\ (i=1,2,3,4,5)$$

となる.

$M=6$ のばあい. p_i は前出と同じとする

このとき

$$\hat{q}_1=2,\ \hat{q}_2=3,\ \hat{q}_3=3,\ \hat{q}_4=3,\ \hat{q}_5=4,\ \hat{q}_6=5$$

となる.

$\bar{q}(\min)$ がえられるばあいの q_i と比較すると, 今度も

$$\hat{q}_i\geq q_i\ (i=1,2,3,4,5,6)$$

となる.

$M=7$ のばあい. p_i は前出と同じとする

このとき

$$\hat{q}_1=2,\ \hat{q}_2=3,\ \hat{q}_3=3,\ \hat{q}_4=3,\ \hat{q}_5=4,\ \hat{q}_6=4,\ \hat{q}_7=5$$

となる.

$\bar{q}(\min)$ がえられるばあいの q_i と比較すると, 今度も

$$\hat{q}_i\geq q_i\ (i=1,2,3,4,5,6,7)$$

となる.

以上の数値例から，このような性質をもつ $\hat{q}_i$ の決定法は，かなり $\bar{q}(\min)$ がえられるばあいの q_i に接近することがわかる．

いいあて法の構造

$\bar{q}(\min)$ がえられる分割形式によってえられる q_i にくらべて $\lfloor 1/p_i \rfloor$ や $\lfloor \lg 1/p_i \rfloor$ から求められる $\hat{q}_i$ が一般に $\hat{q}_i \geq q_i$ であるということは，$\hat{q}_i$ がえられるような質問形式では，“ムダ”な部分が含まれていることを予想させる．

そこで，質問の手順を示すフローチャートをあらためて描いてその性質をチェックしてみよう．

例えば，$M=4$ で（p_i は前出）；等分法，順序法での手順と，$\hat{q}_i=\lfloor \lg 1/p_i \rfloor$ で表わされるばあいの手順をならべてみる（図 11）．

ただしこれらフローチャートで質問の内容は省略した．これを樹木図とよぶ．

これをみると，等分法や順序法では，質問（・印で示す）のあとは，必ず枝路は 2 つに分岐しており，いずれも質問に“ムダ”がないことがわかる．いっぽう，これに対して $\lfloor \lg 1/p_i \rfloor$ 法では，質問してもそのあと 2 つの枝路に分岐しない経路がいくつか含むことになり，これらが質問の“ムダ”に該当する．

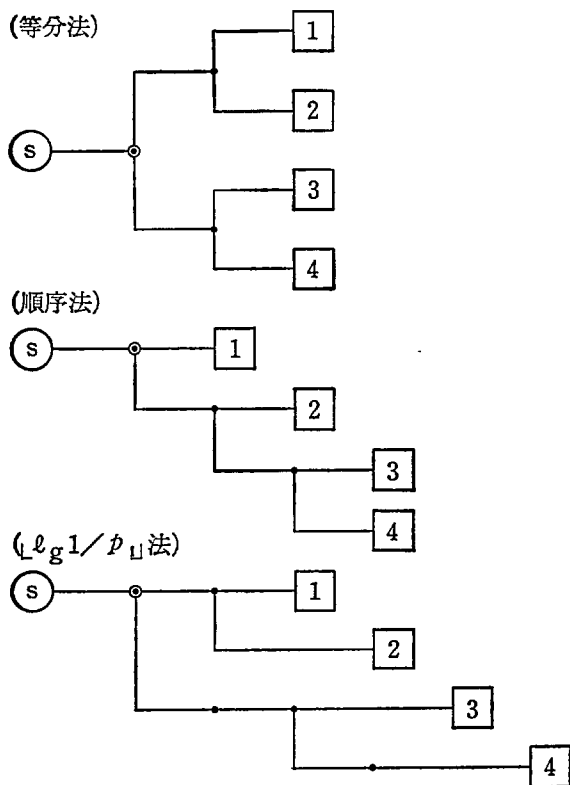
なお，どの手順を採用するにせよ，それらを表わす樹木図について，q_i は最初の質問（・印で示す）のあと，数字札に到達するまでの線分の本数に一致する．

一般に，このような樹木図で表わしたばあい，質問回数 q_i については

$$\sum_{i=1}^{X} 2^{-q_i} \leq 1$$

が成り立つ．

ここで不等号 $<$ は質問にムダがあるとき，そして等号 $=$ は，ムダが

図 11　質問の樹木図（M=4）

ないときに該当する．

数値例：　前出の $M=4$ のばあいについて．等分法では，

$$\sum_{i=1}^{4} 2^{-q_i}=\frac{1}{2^2}+\frac{1}{2^2}+\frac{1}{2^2}+\frac{1}{2^2}=\frac{4}{4}=1$$

また，順序法でも

$$\sum_{i=1}^{4} 2^{-q_i}=\frac{1}{2^1}+\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^3}=1$$

となり，たしかに“ムダ”がないことを示唆する．そして

$\lfloor \lg 1/p_i \rfloor$ 法では

$$\sum_{i=1}^{4} 2^{-q_i}=\frac{1}{2^2}+\frac{1}{2^2}+\frac{1}{2^2}+\frac{1}{2^4}=\frac{11}{16}<1$$

となり，“ムダ”のあることを示している．

さて，質問にムダがないときの判定式

$$\sum_{i=1}^{M} 2^{-q_i}=1$$

これに発生確率の条件式

$$\sum_{i=1}^{M} p_i=1$$

を組合わせる．つまり，左辺どうしが等しいとおき，さらに，ここで $i=1, 2, \cdots\cdots$ の各項について，それぞれ等しい，つまり

$$2^{-q_i}=p_i \qquad (i=1, 2, \cdots\cdots)$$

とおけるなら，

$$q_i=\lg\frac{1}{p_i}$$

なので

$$\bar{q}(\min)=\sum p_i q_i=\sum p_i \lg\frac{1}{p_i}$$

と求められるはずである．

ただし q_i は回数で整数だから p_i はそのような条件をみたすときに限る．

しかし一般には

$$2^{-q_i} \fallingdotseq p_i$$

としなければならない．このとき，

$$q_i \fallingdotseq \lg\frac{1}{p_i}$$

このばあいは p_i を既知としても q_i を導くことはできない．このとき

$$\sum p_i \lg\frac{1}{p_i} \equiv \bar{q}'(\min)$$

とおく.

ここで p_i が小さくなれば $\lg\frac{1}{p_i}$ は急増し $\lfloor\lg\frac{1}{p_i}\rfloor$ なる整数に接近する.

他方, M が大になれば, 一般に p_i が小さくなる傾向にあるから, このとき $\bar{q}'(\min)$ は $\bar{q}(\min)$ に接近することになる.

数値例: 前出の $M=4, 5, 6, 7$ のばあいについて $\bar{q}'(\min), \bar{q}(\min)$ そして両者の相対的なズレ $\eta_M \equiv \{\bar{q}(\min)-\bar{q}'(\min)\}/\bar{q}'(\min)$ も求めてみた(表 8).

これより

$$\eta_3 > \eta_4 > \eta_5 > \eta_6 > \eta_7$$

であることがわかる.

この傾向は, M が大きくなれば $\bar{q}'(\min)$ が $\bar{q}(\min)$ に接近するという, さきに予想した傾向を裏づけている.

$\bar{q}(\min)$ をうるための分割形式

(0)　第0次接近

$\sum q_i$ を最小にすることで満足する. そのためには等分法を, それができないときは等準法を採用すればよい.

(1)　第1次接近

p_i が大になるほど q_i が小になるような分割形式を採用する. 具体的には

$$q_i = \lfloor\lg 1/p_i\rfloor$$

となるように選ぶ.

この程度では, まだ質問にムダがあることは, すでに $M=4$ のばあいでみてきた. つまり $\bar{q}(\min)$ がえられる保証はない.

(2)　第2次接近

表 8　発生確率と質問回数

$(M=4)$

i	1	2	3	5
$p_i \lg 1/p_i$	.5287	.5210	.4643	.3321

$p'(\mathrm{min})=1.85$
$q(\mathrm{miu})=\Sigma p_i q_i=19/10=1\ 90$
$\eta_4\equiv q-q'/q'=0.0270$

$(M=5)$

i	1	2	3	4	5
$p_i \lg 1/p_i$	.5282	.5086	.4643	3.874	.2600

$q'(\mathrm{min})=2.15$
$q(\mathrm{min}=\Sigma p_i q_i=33/15=2.20$
$\eta_5=0.0227$

$(M=6)$

i	1	2	3	4	5	6
$p_i \lg 1/p_i$	.5164	.4928	.4557	.4012	.3226	.2091

$q(\mathrm{min})=2.40$
$q(\mathrm{min})=51/21=2.43$
$\eta_6=0.0125$

$(M=7)$

i	1	2	3	4	5	6	7
$p_i \lg 1/p_i$	.5000	.4760	.4442	.4012	.3450	.2721	.1716

$q'(\mathrm{min})=2.61$
$q(\mathrm{min})=74/28=2.64$
$\eta_7=0.0115$

文字の発生が等確率なら $\sum q_i$ を最小にする等分法，それができなければ等準法を採用して $\bar{q}(\mathrm{min})$ がえられた．そこで，等確率発生でないばあいでも，p_i をいくつかまとめたサブグループどうしを，同じ確率になるように組合わせて，そのあと第0次接近を適用する．

くわしくは：——最初の2分割は，各サブグループの発生確率を0.5に近くなるよう調整し，つぎの分割は，その各サブグループを再び，この半

分の 2.05 づつになるように調整する．

具体的な手順：　前出の $M=7$ のばあいに適用してみる．

数字札1から7を2分するのに，まず前半と後半の2つサブグループにわけ，それぞれのグループの確率の和（合成確率）が等しくなるようにする．そのためには，確率の大きいほうから累和を求めておいて利用する．つまり，累和が 0.5 近いところで2分する．2分されたサブグループは，それぞれ再び同じ要領で2分する．つまり前半のサブグループは確率の累和 0.25 付近で，後半のサブグループでは 0.75 付近で分割する．以下，それぞれのサブグループをまた2分していく．こうして確率の累和の切れ目は，前半のグループで 0.25±0.125，つまり 0.125 および 0.375 となり，後半のグループで 0.75±0.125，つまり 0.625 および 0.875 となる．

この方針にもとずいて実施した分割と，質問の手順を示す（図12）．

この手順は，$\bar{q}(\min)$ をうるときの手順に一致している．

なお，この分割形式を**分確法**とよぶ．

（3）　第3次接近

これは，第2次接近の改良した手法である．

第2次接近では，字種の全体を等確率に近い2分割することから始めて，等確率をめざしつつ細分割を続けていくのに対し，第3次接近では，逆に，小さいほうの発生確率から等確率に近くまとめて，たえず等確率をめざしながら全体をとりこんでいくという手法である．

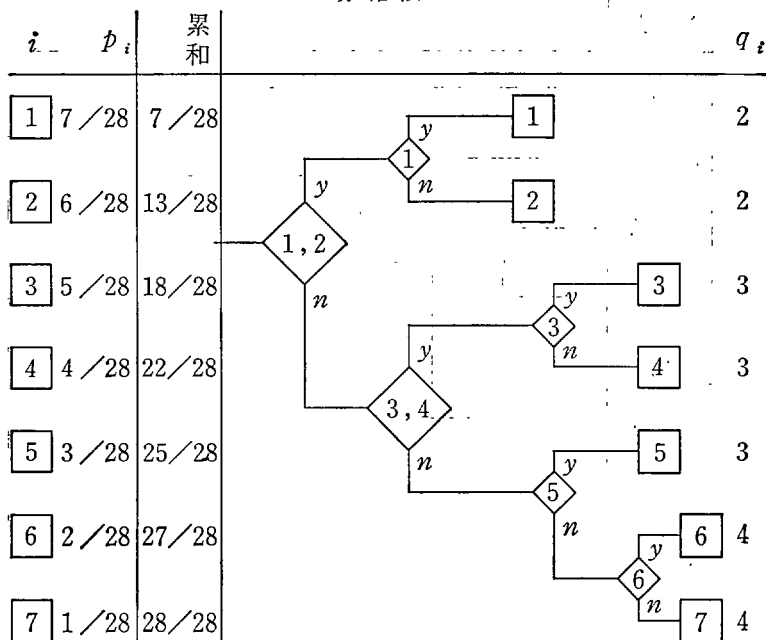
具体的な手順：　前出の $M=7$ のばあいに適用してみる．

①数字札を確率の大きいほうから順にならべる．

②確率の小さいほうから2コをまとめ1枚の札とみなし合成確率を求め，その結果の数値を使って再び大きさの順序にならべかえる．それを合成確率が1になるまで繰返し続ける．

③合成確率1に近いところから，2分された番号のサブグループの一方

図 12　質問の手順（$M=7$）
分確法

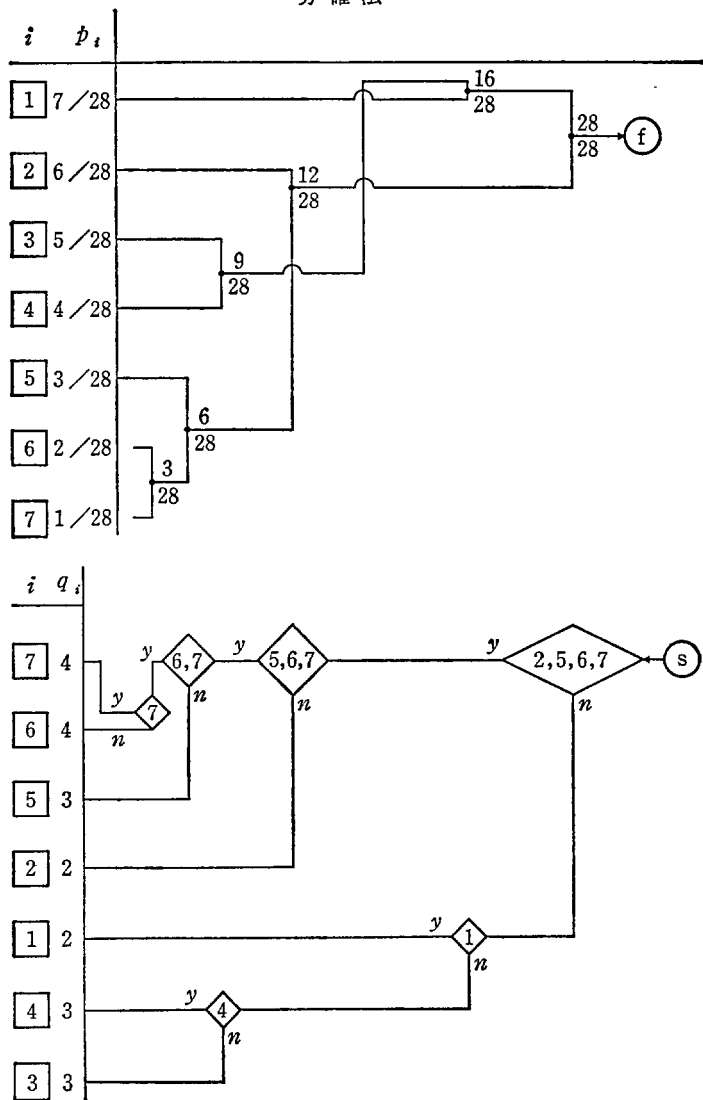
i	p_i	累和	q_i
1	7／28	7／28	2
2	6／28	13／28	2
3	5／28	18／28	3
4	4／28	22／28	3
5	3／28	25／28	3
6	2／28	27／28	4
7	1／28	28／28	4

——ここでは線の下に属するほう——から質問していく．yes なら2分した下の線を1ステップ逆にたどる．no なら上の線を1ステップたどる（図12)．この分割形式を**合確法**とよぶ．

この例（$M=7$）で与えられた $\bar{q}_i$ の系列では，第2次接近の分確法によっても第3次接近の合確法によっても，分割形式が異るだけで質問回数 q_i は一致する．

なお，この第3次接近による合確法は，常に $\bar{q}(\min)$ がえられる分割形式であることが“数学的帰納法”によって証明されている．

（事例1）ヒト（日本人）の血液型をいいあてるばあい．

A, B, O 式の血液型分類によると，日本人は O 型が 31.5%．A 型が

図 13　質問の手順（$M=7$）
分 確 法

37.3%．B 型が 22.1%．AB 型が 9.1% というデータがある．古畑種基『血液型の話』岩波新書　1962.

これは $M=4$ のばあいにあたる．いま%の大きいほうからつまり A, O, B, AB とならべかえてコード番号をつけて発生確率を簡単に

$$p_1=4/10,\ p_2=3/10,\ p_3=2/10,\ p_4=1/10$$

とおけば，前出した順序法の例と一致して $\bar{q}(\min)=1.9$ がえられる．これはたしかに等分法のばあい $\bar{q}=2.0$ よりは小さくなるが，その相違の程度はわずかである．

（事例 2）　氏名をすべて「カナ」で表わしたばあい

「カナ」イロハ 48 文字――のうち，ヰ，ヱは現代かなづかいでは使われない．またヲは助詞として使うだけなので，これらを除くと 45 文字となる．

「カナ」の発生確率の調査は，伊藤栄一氏が第一生命で，保険に加入している人の氏名について行われたものによる（表 9）．

これを使い，第 2 次接近（分解法）で $\bar{q}$ を求めると 5.11 となる．

第 3 次接近（合確法）で求めると 5.09 となる．

なお，$\bar{q}_{mm}=\Sigma_{pi}\lg\frac{1}{p_i}=5.05$ でありちなみに $\lg 45=5.58$ である．

4. 情報の量の定義と基準化

（A）　出にくさの指標

実験者が，つぎつぎと数字や文字をもち出してきて被験者にいいあてさせるという状況は，被験者からみれば，実験者は“情報源”に相当する．そこで一般に，情報源から発生する文字をいいあてるための平均質問回数 $\bar{q}$ を，文字がもつ“知りにくさ”からみた平均情報量（エントロピー）と考える．ただし $\bar{q}$ の最小値 $\bar{q}(\min)$ を念頭におくことにするので，こ

表 9 「カナ」の出現率（伊藤栄一氏のデータ）

カナ	出現率%	累和%	$-p \lg p$	カナ	出現率%	累和%	$-p \lg p$
シ	8.5	8.5	.302293	ヒ	1.5	86.2	.090883
カ	6.3	14.8	.251276	モ	1.3	87.9	.081449
チ	5.0	19.8	.216096	ノ	1.3	88.8	.081449
タ	4.7	24.5	.207326	エ	1.2	60.0	.076570
ウ	4.6	29.1	.204342	ロ	1.2	91.2	.076570
マ	4.5	33.6	.201327	ケ	1.1	92.3	.071570
ヨ	4.0	37.6	.185754	ソ	1.0	93.3	.066439
イ	4.0	41.6	.185054	ユ	0.9	94.2	.061163
ク	3.8	45.4	.179279	ニ	0.8	95.0	.055726
オ	3.8	49.2	.179279	セ	0.8	95.8	.055726
キ	3.3	52.5	.162406	ホ	0.7	96.5	.050109
ミ	3.2	55.7	.158905	ル	0.6	97.1	.044285
サ	3.1	58.8	.155359	ム	0.5	97.6	.038219
コ	3.1	61.9	.155359	テ	0.5	98.1	.038219
ト	2.9	64.8	.148126	ノ	0.5	98.9	.038219
ン	2.7	67.5	.140694	メ	0.4	99.0	.031863
ヤ	2.4	69.9	.129140	ヘ	0.4	99.4	.031863
ナ	2.3	72.7	.125171	ネ	0.4	99.8	.031363
ツ	2.1	74.3	.117043	レ	0.1	99.9	.009966
ハ	2.0	76.3	.112877	ヌ	0.1	100.0	.009966
ワ	1.9	78.2	.108639			Σ	5.054852
ラ	1.7	79.9	.099931				
ス	1.6	81.5	.095453				
ア	1.6	83.1	.095453				
フ	1.6	84.7	.095453				

れの基準値として $\sum p_i \lg 1/p_i$ を採用することにする.

情報源から発生する文字列で，各文字の発生確率は，その文字の“出やすさ”の程度をあらわす．つまり発生確率が大きいほど，出やすさが大きいということになる．今度は，“出にくさ”の指標を考える.

（1） 発生確率の“逆数”を採用するばあい.

発生確率 p の範囲は

$$0 \leq p \leq 1$$

これに対応して，逆数 $1/p$ は

$$\infty \geq \frac{1}{p} \geq 1$$

となる．このとき出にくさの指標は 1 から ∞ までの範囲にひろがる．そこで，この範囲を“縮退”させるために対数をとる．対数を底を 10, e, 2 にとったときをそれぞれ単位（decit), (nat), (bit) をつけて“出にくさからみた情報量”とする.

文字列の長さを n としたとき.

総情報量 I は

$$I = \sum_{i=1}^{M} n_i \lg \frac{1}{p_i} \quad \text{(bit)}$$

ここで n_i は i 番目の字種の文字列の中のコ数，また $\sum_{i=1}^{M} n_i = n$ である.

これより 1 文字当りの情報量，つまり平均情報量（エントロピー）H は

$$H = \frac{1}{n} I = \sum_{i=1}^{M} p_i \lg 1/p_i$$

となる.

こうして，エントロピーは，前出の“いいあてにくさ，知りにくさ”からみた平均情報量——つまり平均質問回数の最小値——と関連づけられる.

“出にくさ”で定義した情報量

一般に，情報は

① いつ役に立つかわのらない情報（データ）

② そのうち役に立つ情報（知識）

③ いますぐ役に立つ情報（情報——狭義の——）

にわけられるとする見方がある.

この見方は情報は“役に立つ”という観点から価値づける態度を反映しており，“すぐに”役立つのが最高の価値というイメージで，狭義の情報とよばれているようである.

この“役に立つかどうか”という観点は，たしかに，情報の価値を計量する1つの観点ではあるが，きわめて人間的であり，しかも“主観的”な立場での評価であるといわねばならない.

このほかにも，情報を評価する観点ないし立場はいろいろ考えられよう.“役に立つ”観点は，ある“目的達成”のためにという前提が条件となっている. その目的は種々あっても基本的に共通しているのは目的志向であり，そこには“打算”と“理性”が介在する.

他方，同じく人間的ではあるが，もっと根元的な“生きること”と直接にかかわる“直感的”で“生理的”ともいえる情報の価値づけ，そしてそれを計量化する方法もある.

それは，A，トフラーのいう「目新しさ」の観点である. A. トフラー『未来の衝撃』実業の日本社　1975.

さきの“出にくさ”の指標をみたときの情報量は，この“目新しさ”の指標として採用できる.

これの観点は，1コの文字や数字などだけでなく，まとまった文字列（用語，語い）さらには，概念などを評価する際にも適用できる.

① 事実現象に関する情報の例

ⓐ 犬が人をかんでもニュースにならないが人が犬をかめばニュースに

なるというニュースの意味についての事例説明は，ニュースバリューというものが，事象の発生確率と関連していることを想像させる．つまり，発生確率が小さいと思われている事象ほど，その事象が発生したとき，ニュース性は高いと判断されていることを予想させるからである．

ⓑ 狼少年のはなし．「狼がくる」と最初に伝えられると，あわてて被害を受けないような対策や行動を選択する．影響を与える効果は大きいから情報量は大きい．しかし，実際には“来なかった”ことが度重なると“またか”“きたためしがない”としてもはや何の対策も講じないようになっていく．行動の面でも精神面でも，これまでと違う変化があらわれなくなってしまう．情報としての効果は失われてしまったことになる．

たとえ聞く度ごとに狼が実際やってくるばあいも，対策は万全になっていき，最初に襲われたときのような恐怖感や不安感は失われて冷静に対処するようになっていく．やはり情報量は減っていくとみてよい．

② 意識現象に関する情報の例

「私，アナタが大好き」というコトバを若い女性から初めて聞かされるとき大感激するだろうが，その女性と結婚して一緒に生活するようになってもあいかわらずいい続けられると，また始まった．いい加減うんざりだと受止めるようになり，やがては聞いても聞かぬふりをするようになってしまう．同じセリフが同じ相手から繰返し聞かされると情報としての価値は減少する．そして相手が違うとなると，同じセリフが再び新鮮さをもつようになり価値は高まる．

このように，同じコトバ（セリフ）でも聞き慣れてしまうと，それがもつ情報量はへっていくとみている．これは，情報が，事実現象，情報現象，意識現象いずれについて記述するのに使われたばあいでも，それが出現するかどうか，あるいは出現する度合いにかかわらせて評価する立場をとることを意味する．

つまり，情報が表わす現象の“稀少性”に価値をおく評価法で，これの程度の指標として情報量が採用できることがわかる．

渡辺紳一郎氏は．「切手の値段」というエッセイの冒頭に，経済学概論で習うことだが，「価値」を決定する1つの原因として「稀少性」ということがある……これを純粋な形で示しているのは蒐集家における古郵便切手の値段である．美しい絵だとか，立派な刷りだとかによって値がきまるものではない．完全に稀少性のみによって決まるのである．……と述べている．渡辺紳一郎『絵本　博物誌』四季社　1958

各国語のエントロピー

文字列における各文字の出現が，字種を M として，

（0）　等確率とするときを0次近似のエンニロピーとよび H_0 で表わす

$$H_0=\sum_{i=1}^{M}p_i\lg\frac{1}{p_i}=\lg M$$

（1）　各文字が独立に出現するとして，各文字の出現度数を出現確率とみたときを1次近似のエントロピーとよび H_1 で表わすと

$$H_1=\sum_{i=1}^{M}p_i\lg\frac{1}{p_i}$$

となる．

これまでの例では，このような出現確率を念頭において情報量やエントロピーを扱ってきたが，文字列の中の文字を単独でなく，2連字，3連字，……をひとまとめにして，それらが独立に出現するととらえたばあいのエントロピーも考えられる．

（2）　2連字が独立に出現するとしたときを2次近似のエントロピーとよび H_2 で表わすと

$$H_2=\frac{1}{2}\sum_{i=1}^{M^2}p_i^{(2)}\lg\frac{1}{p_i^{(2)}}$$

ここで $p_i^{(2)}$ は2連字の i 番目の出現確率である．

（3） 3連字をとったとき，3次近似のエントロピーとよび，H_3 で表わすと，

$$H_3=\frac{1}{3}\sum_{i=1}^{M^3}p_i^{(3)}\lg\frac{1}{p_i^{(3)}}$$

ここで $p_i^{(3)}$ は3連字の i 番目の出現確率である．

さらに，n 連字についての n 次近似のエントロピーも，形式的には，上記の拡張として数式表現はできるものの，$n=4$，つまり4次以上になると，その連字の出現度数をあらためて調査し直さねばならず厄介になってくる．

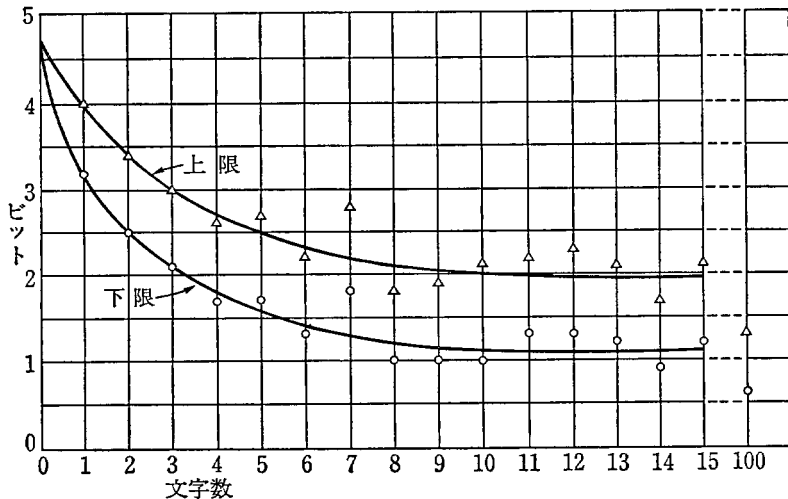
（4） 高次エントロピー

$M=27$（英語）では，4連字の異なる字種は $M^4 \fallingdotseq 5.3\times10^5$（≈50万），5連字となると $M^5 \fallingdotseq 1.4\times10^7$（1,400万）となる．

$M=54$（日本語）では，4連字で $M^4 \fallingdotseq 8.5\times10^6$（850万），5連字で $M^5 \fallingdotseq 4.6\times10^8$（≈5億）に達する．

C. E. Shannon (1949) は，実験的な方法で H_∞ を求めることを考えた．これは，1種の「いいあて法」である．

英文章を始めの文字から1字づつ「この文字ですか」といいあてさせる．正しければ yes と答え，誤りであれば no と答え，いいあてるまで繰返させる．こうしてすべての文字について問いの回数を全文字数で割ってエントロピーとする．実際に行われたのは文字数 $n=102$ で，質問の回数が198であったから，エントロピーは 198/102=1.94 (bit/文字) となった．ただし，つぎの文字が明白なばあいも質問回数に数えるので，真の値は，これより小さいと見積られる．上記の数値は“上限値”である．

図 14　英字エントロピー
Shannon, C. E. Bell System Tech, J. 30 50 ('51)

それは，被験者が自国語について，①使用回数や②相互関係を直観的に知っていると仮定しているのでふつうの人間が不完全な言語知識しかもたないことを考えると，質問回数はその分だけ増加する傾向があるからである．

C. E. Shannon は，その後（1951）高次近似のエントロピーを間接的に推定する「予想ゲーム法」も考察した．

ある小説から 15 文字からなる系列を 100 組ランダムにとり出した文字列についてこの法を適用してエントロピーの上限値および下限値を実験的に求めている．またこの結果を外挿して H_{100} の上限値を 1.3．下限値を 0.6（bit/文字）と推定している（図 14）．

露語では $M=32$ で $H_8 \approx 1.36$ がえられた．近似を高めていくと H_{15} で安定しそれ以後はほとんど小さくならないことがわかった．これより H_∞

＝1.0～1.2（bit/文字）と推定された．

N．バートンとJ．リクライダーはShannonの法で実験して，ある文字がでたあと1000番目にどんな文字がでるかは予言しにくい．そして文字相互の結びつきは約30字の範囲までであると推定している．

これは一般の欧語についても成立しよう．

ア．エヌ．コルモゴロフらは，シャノンの「いいあて法」を改良した．

これを，あまり知られていないゴンチャロフの作品から取出された文字列に適用した．

この作品は①平均的な散文である，②読むのに特に困難ではなく，特殊な言いまわしもない，ということで選ばれた．これより露語でも $H_\infty \fallingdotseq 1$ と推定された．

以上の考察から各国語のエントロピーについて，一般的にいえることは，

H_0 はアルファベットの字種 M が大きくなるほど大きくる．

H_N は N が大きくなるほど小さくなっていく傾向がある．一般に

$$H_0 \geq H_1 \geq H_2 \cdots\cdots \quad \geq 0$$

このような傾向をもつことは，先行文字について知るほど，つぎの文字はいいあてやすくなることからなっとくできる．

H_∞ は計測することは困難である．正確な値は，どの国語についてもわからない．しかし十分な確信をもって，主要ヨーロッパ語では，ほぼ一致する値，約1.0であるといえそうである．

連字について

文字 Q で始まる単語を，三省堂編『初級コンサイス英和辞典』で引くと

quack（あひるがガーガー鳴く声，いんちきな）に始まり，quste（引用する）に終る．36コの単語のすべてがqのあとにuが続く．語頭でな

くても squirral (りす), conquer (征服する) など，やはり q のあとに u が続く.

語頭となる文字のつぎにくる文字の分布は，ふつうの辞書で調べればわかる.

そして，語尾の文字の前にくる文字の分布は逆引辞典で調べられる.

これは，単語の末尾をそろえ，末尾から逐次アルファベット順にならべた辞典で，例えば，郡司『英語学習逆引辞典』開文社によると，——g で終る単語は 514 コ，この前の文の文字の内訳は，a が 12, e が 3, g が 1, i が 8, n が 465, o が 11, r が 1, u が 13 となっており，n が群を抜いている．だから n がくれば a と続く確率は 90% である.

文字の連接は 2 文字にとどまらず，あとの文字にも波及する．qu のつぎは必ず「母音」である．th のつぎは e がくる確率が高く l や x は続かない.

文の種類とエントロピー

これまで文字列は，一般的なふつう文から取出して考えてきた．しかし文の種類（ジャンル）が異なればエントロピーに平均的な差異が生じてくると予想される.

（i） ふつう文と技術文

ふつう文のエントロピーを H_f, 技術文のエントロピーを H_t とすると，一般に

$$H_f > H_t$$

と予想される.

技術文では ①語いの種数が少ないとか，特殊な通用語や表現が多いからというのがその理由とされる.

どの国の言語であれ，専門の技術文のほうが芸術，文芸文より“はるか

に”読みやすいのである.

ただし，技術文も特殊になると例外がある.

① 特殊用語が，高密度の意味内容(学術用のコトバ——例えば現代数理論学で使われている記号的なコトバ——）をもたされて使われるばあいや.

② 特殊用語が，ふつうの人に解読されにくいかうくようする（泥棒の使う陰語——ごく短かいコトバだが非常に多くの意味をもつ）ことを目的として使われるばあいなど.

（ii） 文学的な文章

① 文章型が異なるとエントロピーも異なる.（文学作品のジャンルが異なれば，それぞれ特有な表現法が採用されるから）

② 1つの文学作品でも，部分的にはエントロピーが異る.（対話文と描写文など）

③ 筆者が異なればエントロピーは異なる，よく知られた型どおりの記述——文学作品では自然描写の部分——ではエントロピーは小さくなる.

これに対し，意外な内容に富み，型通りでないコトバが使われるならエントロピーは大きくなる.

特にエントロピーが大きくするには，内容を複雑にし，訳がわからない文章にすればよい.

④ 小説と詩では，詩のほうがコトバに補則的な制限を課するので，つづく文字や単語をいいあてやすく，それだけエントロピーは小さくなる.

しかし他方，詩は，月並みでない内容を盛りこませるので，結果的には散文と同じ程度のエントロピーにもどるという.

話しコトバのエントロピー

これまでは，文章，つまり書きコトバを前提にして，そのコトバが意味をもつばあいを考えてきた.

会話や話しコトバでは，話す際の抑揚（イントネーション）など意味的でない部分も含まれており，これらも情報をもたらすと考えねばならない．

これを手がかりにして，知人の声を聞きわけたり，話し振りからその人の性格や気質を知ることもできるからである．これらを「意味的でない情報」とよび，話しコトバにうつしとれる「意味上の情報」と区別すると，

両者が混在する一般の会話で，両者が“矛盾”していたとすると，人は，前者を信頼するといわれている．

直接，相手と面と向っての会話では，声以外に，顔色，表情や態度からも情報が伝わってくる．「目は口ほどに物をいう」のである．

いま，意味的でない情報量を H_A，意味上の情報量を H_B とすると，この比 H_A/H_B は，話す速度に依存することが知られている．

（a） 非常に速いばあい：声の聞きわけがしにくく，抑揚のニュアンスが追跡できなくなる．それは「意味」を理解するのが精いっぱいで「意味ではない」部分の解読まで処理が及ばないためで，このときは H_A が小さくなるので $H_A/H_B \fallingdotseq 30\%$ と小さい．

（b） ふつうのばあい：$H_A/H_B=75\%$ と大きくなり，

（c） ゆっくりのばあい：$H_A/H_B=130\%$ に達するといわれている．

文字と単語のエントロピー

文章の総情報量 I は，文字のコ数を n，文字のエントロピーを H_l として $I=nH_l$ いっぽう単語のコ数を m，単語のエントロピーを H_w とすれば $I=mH_w$

ここで $s\equiv n/m$ は，1単語当りの平均文字数で，英語では 4.1～4.5 字，露語で 5.3 字，独語で 5.9 字，ポリネシア語で 3.2 字と求められている．

音楽における音名のエントロピー

音楽情報は，①メロディ　②リズム　③ハーモニィの3つの要素から成立つとされる．

①は「ふし」で，音の高さの変化が主な印象となるが，音の長短や強弱も関与する．

②は音の長短の“律動”と，音の強弱の“拍節（拍子）”を取出したもので音楽の進行に“力動さ”や“釣合さ„を与える．

③は“協和”のことで，音楽に“多彩な変化”を与える．

このうち，メロディの音の高さだが，これはデタラメにならべられるものでなく，音楽的な訓練をうけると，未知のメロディでもつぎの音の高さが予想できるといわれる．

いま音の高さを音名で b, $c^\sharp$, d, e, F^*, G, $G^\sharp$, A, B, $C^\sharp$, D, E, の12種について，エントロピーを求めると

$$H_0 = \lg 12 = 3.58 \text{ bit}$$

$$H_1 = -\Sigma\, p_i \lg p_i = 3.43 \text{ bit}$$

がえられている．

（B）　相対エントロピーと冗長度

各国の言語のエントロピーを比較するばあい，字種 M が異なれば，0次近似のエントロピー H_0 からすでに異なるので，実際のエントロピーどうしを直接くらべるのは適切ではない．

そこで，それぞれの言語でエントロピーが最大値となる H_0 で基準化した**相対エントロピー** H_r を採用する．つまり

$$H_r \equiv H/H_0$$

ここで H_r の範囲は

$$1 \geq H_r \geq 0$$

となる．そして

$$R \equiv 1 - H_r$$

で定義される量を**冗長度**とよぶ.

ここで H_r に注目して文章の特性を指摘してみよう.

(i)　$H_r=1$ のばあい：文章を構成する各文字が独立で等確率発生の状況を意味すち．あいさつのときのような，よく使われるいいまわしなど一切含まれないばあいである．それどころか，異なる文字列がそれぞれ異なる意味をもつことを想定しなければならない．つまり文中の1文字が誤って別の文字に書かれると，全く別の意味をもつことになる.

だから，このような文章では，中の1字が抜けおちたとすると，そこへ入れるべき文字は，アルファベットの字種だけあり，それに応じて異なる意味の文ができあがることになる.

(ii)　$H_r=0$ のばあい：文章で，ある文字のつぎの文字は，その前にでた文字，文字列によって完全にきまってしまうばあいである．文中の1文字が抜けおちても，そこへ入れるべき文字は，自動的，機械的に決まってしまい選択の余地がない．それ以外の文字を入れると文章は意味をなさない.

数学における数式の変形や論理記号式での変形がこれにあたる．つまり数学という学問は，式をたてるまでが情報として有効な領域であり，式の変形（演算や計算）は冗長度 100% の領域であるということになる.

(iii)　実際の H_r $(1>H_r>0)$ のばあい：文中にいくつかの文字が抜けおちていてもそこに入れるべき文字は，数回の試行錯誤を経て，適切なものが見つけられ文の意味が通るようになる．このように適当な冗長度があるのでクロスワードや暗号文などの解読も可能となる.

各国語の相対エントロピー

日本語では $M=54$ として，

$H_0=5.75$, $H_\infty=1.8$ にすれば

$$H_r=1.8/5.75=0.31$$

英語では $M=27$ として

$H_0=4.75$, $H_\infty=1.3$ とすれば

$H_r=0.27$, また $H_\infty=1.0$ とすれば

$H_r=0.21$ となる.

露語では $M=32$ として,

$H_0=5.00$, $H_\infty=1.2$ とすれば,

$H_r=0.24$, また $H_\infty=1.0$ とすれば

$H_r=0.20$ となる.

このように, 日本語, 英語, 露語については, いずれも $H_r≒20〜30\%$. $R≒80〜70\%$ の範囲に収まっているとわかる.

その他の国の言語についてもほゞ同様とみられる.

それは, 人間が“話し”の流れを知覚する“生理学的”な過程での特性と結びついており, 上記のような H_r ないし R が“最適値”なのであろう. 冗長度が大きいようにみうけられるが, 人間の頭脳のはたらきとの関連からみて, この程度の冗長度は“不可欠”なのだといわれている.

5. 情報量からみた情報の伝達

情報の伝達は, 3つのレベルで検討できる.

(i) 文字, 数字など記号列を構成する要素——記号——をいかに“正確に”伝えるか. つまり伝達の技術的側面.

(ii) 記号列が, 伝えたい意味内容を明白に誤らず表現しているか. つまり語義的側面.

(iii) その記号列の意味内容が, 発手の意図した通りに, どこまで受手の意識の変革や行動の選択に有効な影響を与えたか. つまり効果的側面.

人から人へと話しコトバが“ナマ”のまま伝えられていく「耳うちゲーム」を経験してわかるように，最初の発手から出た話しコトバの列が最終の受手に伝わるときには，かなり変化してしまうのが普通である．たとえ，発手と受手の２人だけの情報伝達の場面であっても，その間に介在する通信環境が悪ければ話しコトバは加工，変形をうける．

地下鉄の車内や街の盛り場では，外部から雑音が混入してくるし，話しコトバの一部は洩れたり消失することがおこる．そこで

① 話す速度をおそくして，聞きわけるのに時間をかけても間に合うよう余裕を与える．

② むづかしい語句を使わず，理解するための負担を軽くする．

③ 同じ意味の語句のいいかえを行い，聞きなれたチャンネルが成立するのを助ける．

さらに悪い状況下，あるいは，極度に伝達の正確さが要求されるときは，

④ 語いや単語をスペルに分解して伝えるため「コード用語」を使うなどのくふうがされる．

（A）　コード用語のいろいろ

日本語の 50 音，数字，英字アルファベットのばあい（表 10）

第２次大戦中，米国陸軍が使用したスペル用単語．および国際民間航空機関（ICAO）が開発したスペル用単語（表 11）.

フランス語の ABC 確認用語（表 12）.

スウェーデン語のアルファベットおよび数字（表 13）.

送る文字数 N_0 コ，この各文字を，その文字を含む単語（コード単語）に変えて送るときの総文字数を N, また送る文字の情報量を H_0, コード単語にしたときの情報量を H とすると $N_0H_0=NH$

が成立つから，

表 10　コード用語（日本）

文字	用語
あ	朝日のア
う	上野のウ
お	オワカのオ
き	切手のキ
け	景色のケ
さ	サクラのサ
す	スズメのス
そ	ソロバンのソ
ち	チドリのチ
て	手紙のテ
な	名古屋のナ
ぬ	沼津のヌ
の	野原のノ
ひ	飛行機のヒ
へ	平和のヘ
ま	マッチのマ
む	無線のム
も	モミジのモ
ゆ	弓矢のユ
ら	ラジオのラ
る	留守のル
ろ	ローマのロ
ゐ	井戸のヰ
を	尾張のヲ
い	イロハのイ
え	英語のエ
か	かわせのカ
く	車のク
こ	子供のコ
し	新聞のシ
せ	世界のセ
た	タバコのタ
つ	ツルカメのツ
と	トンボのト
に	日本のニ
ね	ネズミのネ
は	はがきのハ
ふ	富士山のフ
ほ	保険のホ
み	ミカンのミ
め	明治のメ
や	大和のヤ
よ	吉野のヨ
り	リンゴのリ
れ	レンガのレ
わ	ワラビのワ
ゑ	ヱビスのヱ
ん	オシマイのン
a	始めのエー
c	シネマのシー
e	イーストのイー
g	グラムのジー
i	アイスのアイ
k	京都のケー
m	ミルクのエム
o	大阪のオー
q	キューピーのキュー
s	スポーツのエス
u	ユニオンのユー
w	早稲田のダヴリュー
y	ワイシャツのワイ
&	エンドマーク
一	イチ・ヒトツ
三	サン・ミッツ
五	ゴ
七	ナナ（シチといわない）
九	キュウ（クといわない）
百	ヒャク・イチマルマル
b	ビールのビー
d	デパートのデー
f	フランスのエフ
h	ホテルのエイチ
j	ジャパンのジェー
l	レターのエル
n	日本のエン
p	ピースのピー
r	ラジオのアール
t	東京のテー
v	ヴィクターのヴィ
x	クリスマスのエックス
z	終りのゼット
二	フタ（ニといわない）
四	ヨン（シといわない）
六	ロク
八	ハチ
十	トウ・イチマル
〇	ゼロ・マル

表 11 コード用語（アメリカ）

(a) 米国陸軍

Able	Jig	Sugar
Baker	King	Tare
Charlie	Love	Uncle
Dog	Mike	Victor
Easy	Nan	William
Fox	Oboe	X-ray
Geoge	Peter	Yoke
How	Queen	Zebra
Item	Roger	

(b) 国際民間航空機構

Alpha	Golf	Mike	Sierra	Yankee
Bravo	Hotel	Novenber	Tango	Zulu
Charlie	India	Oscar	Uniform	
Delta	Juliet	Papa	Victor	
Echo	Kilo	Quebec	Wiskey	
Foxtrot	Lima	Romeo	Xray	

表 12 コード用語（フランス）

Anatole（アナトール）	Eugène（ユージェヌ）
Berthe（ベルト）	Emile（エミール）
Célestin（セレスタン）	Fançois（フランソワ）
Désiré（デジレ）	Gaston（ガストン）
Henri（アンリ）	Raoul（ラウル）
Irma（イルマ）	Suzanne（シュザンヌ）
Joseph（ジョセフ）	Thérèse（テレーズ）
Kléber（クレベール）	Ursule（ユルスル）
Louis（ルイ）	Victor（ヴィクトール）
Marcel（マルセル）	William（ウィリアム）
Nicolas（ニコラス）	Xavier（ザヴィエ）
Oscar（オスカー）	Yvonne（イヴォンヌ）
Pierre（ピェール）	Zoé（ゾエ）
Quintal（カンタル）	

表 13　コード用語（スエーデン）

A-Adam	K-Kalle	U-Urban	1-Ett
B-Bertil	L-Ludvig	V-Viktor	2-Tvåa
C-Cesar	M-Martin	W-Wilhelm	3-Trea
D-David	N-Niklas	X-Xerxes	4-Fyra
E-Erik	O-Olof	Y-Yngve	5-Femma
F-Filip	P-Petter	Z-Zäta	6-Sexa
G-Gustav	Q-Quintus	Å-Åke	7-Sju
H-Helge	R-Rudolf	Ä-Ärlig	8-Åtta
I-Ivar	S-Sigurd	Ö-Östen	9-Nia
J-Johan	T-Tore		0-Nolla

これより　冗長度 R は，

$$R=1-H/H_0=1-\frac{N_0}{N}$$

数値例

ICAO のばあい，「STOP」の 4 文字は，スペル用単語では「Sierra Tango Oscar Papa」と 20 文字になる．これより

$$R=1-4/20=0.8$$

また，スェーデン語のばあい Abisko の 6 文字は「Adam Bertil Ivar Sigurd Kalle Olof」と 29 文字にする，これより

$$R=1-6/29\doteqdot 0.8$$

いずれも，ふつう文がもつ情報量のレベルまで冗長度を高めていることがわかる．

一般に，誤りが起りやすい状況の下や雑音がある中で，何が何でも情報を伝達したいばあいは，上述のように“冗長度”を故意にふやしている．

もっとも，雑音の中でも意識を集中すれば自分が聞きとりたい情報だけをある程度うまく聞きわけられることがある．これを**カクテルパーティ効果**とよぶ．

逆に，誤りが起りにくい状況や条件の下では冗長度を故意に減らすこと

表 14　航空会社の略語

(Lufthansa Timetable Japan 01 Jul–25 Oct 86 より)

	Airlines Fluggesellschaften
AA	American Airlines
AC	Air Canada
AF	Air France
AH	Air Algéria
AI	Air India
AL	US Air
AM	Aeromexico
AN	Anselt Airlines of Austraralia
AO	Aviaco
AR	Aerolineas Argentinas
AS	Alaska Airlines
AT	Royal Air Maroc
AV	Avianca
AY	Finnair
AZ	Alitalia
BA	British Airways
BD	British Midland Airways
BM	Aero Trasporti Italiani
BR	British Caledonian Airways
BU	Braathens S. A. F. E.
CA	CAAC China
CC	Air Atlanta
CI	China Airlines
CO	Continental Air Lines
CX	Cathay Pacific Airways
CY	Cyprus Airways
DA	Dan Air Services LTD
DL	Delta Air Lines
DU	Roland Air
DW	DLT Deutsche Luftverkehrsgesell-schaft mbH
EA	Eastern Air Lines
EG	Japan Asia Airways
EI	Aer Lingus
EQ	T. A. M. E. Ecuador
ET	Ethiopian Ailines
EU	Ecuatoriana
GA	Garuda Indonesian Airways
GE	Guernsey Airlines
GF	Gulf Air Company
HH	Somali Airlines
HP	America West Airlines
HX	Holiday Express
IA	Iraqi Airways
IB	Iberia
IC	Indian Airlines
IR	Iran Air
IT	Air Inter
JL	Japan Air Lines
JP	Inex Adria Airways
JU	Jugoslovenski Aerotransport
JY	Jersey European Airways
KE	Korean Air Lines
KQ	Kenya Aicways
KU	Kuwait Airways
LA	LAL de Chile
LG	Luxair
LH	Lufthansa
LN	Samahiriya Libyan Arab Airlines
LO	Polskie Linie Lotnicze
LX	Crossair
LZ	Balkan Bulgarian Airlines
MA	Malév Hungarian Airlines
ME	Middle East Airlines
MH	Malaysian Airline System
MS	Egyptair
MX	Compania Mexicana de Aviacion
NS	Nürnberger Flugdienst
NW	Northwerst Orient Airlines
OA	Olympic Airways
OC	Air Cal
OK	Ceskoslovenské Aerolinie
OL	Olt-Ostfriesische Lufttransport GmbH
OS	Austrian Airlines
OZ	Ozark Air Lines
PA	Pan American World Airways
PI	Piedmont Aviation
PK	Pakistan International Airlines
PR	Philippine Airlines

PS　Pacific Southwest Airlines
PU　Pluna
PW　Pacific Western Airlines
QF　Qantas Airways
RA　Royal Nepal Airlines
RC　Republic Airlines
RG　Varig
RJ　Royal Jordanian Airlines
RO　Tarom Ramanian Air Transport
SA　South Afriican Airways
SC　Cruzeiro
SD　Sudan Airways
SK　Scandinavian Airlines
SN　Sabena
SO　Austrian Air Service
SQ　Singapore Airlines
SR　Swissair
SU　Aeroflot
SV　Saudi Arabian Airline
TE　Air New Zealand-International
TG　Thai Airwaysys International
TH　Thai Airways LTD
TN　Trans Australia Airlines
TP　TAP Air Portugal
TR　Transbrasil
TU　Tunis Air
TW　Trans World Airlines
UA　United Airlines
UK　Air U. K.
UL　Air Lanka
UM　Air Zimbabwe
VA　Viasa
VG　Regionalflug GmbH
VO　Tyrolean Airways
VP　VASP
WA　Western Airlines
WO　World Airways
WT　Nigeria Airways

がある．

新聞の3行広告や，不動産業者のアパートやマンションの間取りや設備内容の広告などに特徴的にみられる．

略語，略号は広く一般的に使われている．

航空会社の略語の例（表 14）．

さきの送る文字数 N_0 と，その文字を含むコード単語にした総文字数 N との関係式がここでもそのまま使える．こんどは

① N_0/N: 使われている文字数 N の何割でよいかを表わす

② $(N-N_0)/N\equiv R$: 使われている文字数 N に対して縮められた文字数の割合（消失率）を表わす．

数値例

① 新聞広告における 20 字は，これをふつう文に解読したら 90 字に近くなるといわれる．これより $N_0=20$, $N=90$ とすると

$$R=1-20/90\fallingdotseq0.78$$

これは一般の文章の冗長度 $R=0.8$ にくらべて小さいから，ふつうの文章に“復元可能”なはずと了解される．

② ジュール・ヴェルヌの小説『グランド船長の子どもたち』の中に，船長がビンから取出した“書付け”の文字 250 コのうち 150 コが海水で洗い流されていたという一節がある．このばあい $N_0=250-150=100$ $N=250$ だから $R=1-100/250=0.60$ これも 0.8 より小さいから復元可能である．

③ 航空会社の略語（前出しの表 12）についても，全会社のフルネームをふつう文とみてのべ文字数 $N=1760$，これをすべて略号に直したものを短くした文とみてのべ文字数 $N_0=117\times 2=234$ を使って $R=1-234/1760=0.87$ となる．これは 0.8 を上まわっている．略語による“短縮化”は極度に進んでいる．

一般にふつうの文の短縮は，言語を“符号化”するものであり，新しい簡潔な言語を確立したことにはならない．表現としては短縮され文字数が減少してはいても，頭脳で処理するにあたっては自然言語に翻訳してその意味を理解しているのであるというのがその理由とされている．

（B） 2元符号語による文字の伝送

文字列で表わされた文章を通信回線を使って電気信号（ON-OFF）に直して伝送するばあいを考える．

まず，情報源から発生する文字列は，**情報源符号器** (Source Encorder) により**2元符号**に変換（**符号化＝コーデイング**）させて**通信路**にのせる．受信側では到着した2元符号を**受信復号器** (Source Decorder) に入れ**復号化＝デコーデイング**）してもとの文字列に変換し**受信者**に届ける．

従って，符号化は，あとで復号化が可能であるよう行わなければならないが，さらに符号語を受け終った瞬間に，符号語が終りであることが検知

されるとき（**瞬時**）**復号可能**という．

復号可能な符号語をつくる法

符号化した際の符号語はその特徴により2つのタイプにまずわけられる．

（i） 符号語の長さを一定にそろえる．

アルファベット文字1コを常にある一定の長さ l の符号語に割当てるばあいである．このとき受信側では，長さ l ごとに区切れば復号化できる．

（ii） 符号語の長さを一定にしないばあい．

符号語の長さは符号の樹木図(**符号樹**)を使って割当てるとわかりやすい．始点 s から出る線分を最初の0次節点で2本に分岐させ，上の枝に 0，下の枝に1を割当てる．それぞれの先端（端点）を1次節点として再び2本に分岐させ，前と同様に上，下の枝に0と1の符号を割当てる．以下，これで必要なコ数（字種）の端点ができるまで続ける．変換される符号語は，これら各枝の符号をすべてつなぎならべたものとする．その際，分岐させる枝はすべて同じ次数の節点まで伸ばす必要はない．字種 $M=4$ のばあいについて，いろいろな符号語のセットがつくれることを示す(図15)．

Ⅰでは符号語は長さ 1, 2, 3, 3 である．Ⅱでは長さ 2, 2, 3, 2 である．Ⅲでは長さ 2, 2, 2, 2 で等長符号語になる．Ⅳでは長さ 1, 3, 3, 3 である．以上，ⅠからⅣまでは，すべて端点まで使った符号語なので，すべて復元可能となる．Ⅴでは長さ 1, 2, 2, 1 だが，最初と最後の符号語 0 と 1 は途中の節点で打ちきって符号語をつくっているので復元不可能である．

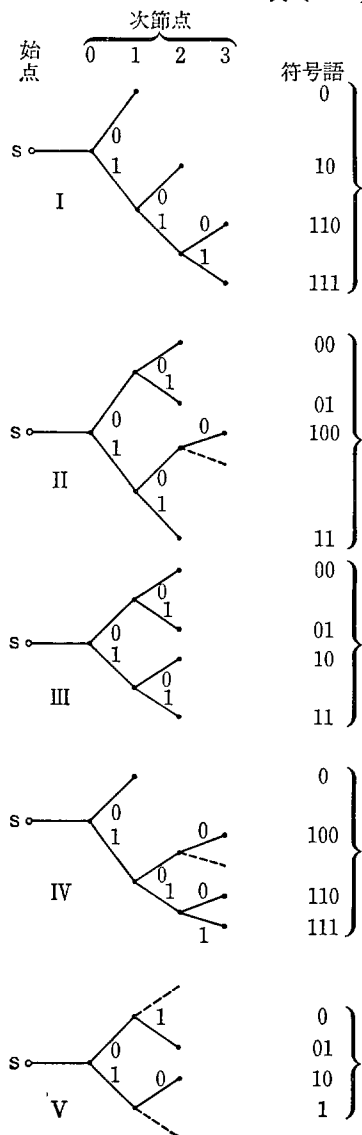
文字 A, B, C, D をこれらの符号語の上から順に割当てたとして，いま符号列

0 1 1 0 1 1 1 0 1 0 0

を受取ったとする．

Ⅰ，Ⅲは端点符号語がペアで使われているから完全に復元可能．ペア符

図 15　符号語のセットの例 (M=4)

0
01
10
1
V

号語で構成される端点符号語は“完全である”という．IIでは復号できない部分がでてくる．それは端点符号語がペアで採用されていないからである．IVもペアで使われていないが，偶然この符号列は，うまく復号できる．Vはどこで区切るかによって別々の文字列に復号されてしまうことなどがわかる．

一般に r 元符号 (r=2 なら符号は 0 と 1) を使って字種 M の文字に対応する符号語をつくるばあいその長さを $l_1, l_2, \cdots\cdots l_M$ とすると，これら符号語が復号可能であるための必要・十分条件は

$$\sum_{i=1}^{M} r^{-l_i} \leq 1$$

ここで等号＝は符号語が“完全”なばあい，不等号 < はそうでないばあいである（符号長に関する Kraft の定理）．

（C） 通信の質の確保

情報源符号器と受信復号器の間の通信路を2元符号語が伝送されるときに，一般に雑音が混入したり，また符号の一部が洩れたり消失したりして符号列が変形をうけることを予想しなければならない．こうした符号列を乱す原因を**雑音**（広義）とよぶ．そこで

（1） 雑音がなければ符号器から符号0が送り出されるならそのまま復号器に符号0が到達し，1が送られるなら1が到達する．

いま，符号器側を $x(0,1)$，復号器側を $y(0,1)$ で表わすと，

$x=0$ で常に $y=0$, $x=1$ で $y=1$ となる．従って受とった符号 $y=0$ なら $x=0$, $y=1$ なら $x=1$ と判定すればよい．

これを条件付確率 $p(y|x)$ で表わすと

$p(0|0)=1$, $p(1|1)=1$　で

$p(1|0)=0$, $p(0|1)=0$　となる．伝達は「確定的」である．

（2） 雑音があると，$x=0$ であっても必しも $y=0$ でなく $y=1$ となることもある．$x=1$ であっても $y=0$ となることもある．

条件付確率で表わすと

$$p(0|0)<1, \quad p(1|1)<1$$

$$p(1|0)>0, \quad p(0|1)>0$$

となる．伝達は「確率的」な性質をもつことになる．

通信を「確定的」にするには，雑音を“完全”になくせばよいが，これは現実には無理なので，一般には，誤りがあったとき，①それに気づかせること（**誤り検知**）ができるか②どこが誤っているか指摘（**誤り訂正**）できるようにくふうする．

このようにして通信の「質」を確保する．

誤り検知または誤り訂正させるというのは，それ自体1つの情報である

からそれだけ余分の情報を追加して送り届けねばならない．ここでは情報は2元符号で送られているから結果的にはこのために符号語をさらに付加することを意味する．

誤り検知や訂正の機能をもたせるために送信側に**通信路符号器**（Channel Encorder）をまた受信側に**通信路復号器**（Channel Decorder）を付加する．

（i）誤り検知

文字 A, B, C, D の字種 $M=4$ からなる文字列を情報源符号器により等長の2元符号語 00, 01, 10, 11 に変えたとして，そのあと通信路符号器で，00 には 0，01 には 1，10 には 1，11 には 0 の1ビットを追加する．もとの2ビットを**情報ビット**，追加の1ビットを**チェックビット**，両者あわせて**送り出しビット**とよぶ．

チェックビットを0にするか1にするかは，送り出しビットに含まれる“1”のコ数を合計したとき偶数（0を含む）になるように決める．これを**偶数パリティ**とよぶ．

これより，受信側の通信路復号器が，到着した符号語の3ビットのなかかの“1”の合計の奇，偶を調べる．調べるには奇偶カウンターを使う．そして偶数ならOK！とする．

これが**パリティチェック法**である．ただし到着した符号語に誤りが2ビットあるとこの方式では見外すことになるが，この事態が発生する確率は実用上は無視できるほど小さい．

実際には，符号語は十分長いから，これを k コづつに区切りブロック化し，そのあと1ビットを追加して偶数パリティにして送り出している．このとき送り出しビットの伸びは $(k+1)/k$ となる．

（ii）誤り訂正

前出の字種 $M=4$ のばあいで考える．今度は通信路符号器で 00 には

000, 01 には 110, 10 には 101, 11 には 011 と検知のばあいよりチェックビットを2ビット追加して送り出しビットとする．ここでチェックビットを0にするか1にするかは，情報ビットから始まる5ビットを x_1, x_2, x_3, x_4, x_5 と記号化したとき，チェックビットの1番目 x_3 は x_1, x_2, x_3 のブロック（ブロック S_1）が偶数パリティになるように，以下，x_4 は x_2, x_4（ブロック S_2）が，x_5 は x_1, x_5（ブロック S_3）が，それぞれ偶数パリティになるようにする．

さて，受信側での訂正の原理と方法はつぎの通り．

受信側でブロック S_1, S_2, S_3 にそれぞれ対応させて奇偶カウンターをおく．誤りなしに届くとどのカウンターも常に“0”を表示するが，例えば x_1 が誤まると S_1 と S_3 が“1”を示す．これを 101 で表わせば，以下，x_2, x_3, x_4, x_5 がそれぞれ誤ると 110, 100, 010, 001 と異なる3ケタの符号語（症候群）で表示されることになる．

さて，奇偶カウンター3コなら症候群の異なる符号語は $2^3=8$ 種だがこのうち 000 は“誤りなし”なので除けば，送り出し符号語にして7ビットまでの訂正が可能なはずである．

一般に，情報ビット，送り出しビットの長さをそれぞれ L, N とすると，誤り訂正ができるめには $2^{(N-L)}-1\geq N$ でなければならない．このような符号語を **Code (*N*, *L*)** で表わす．

訂正符号との対応をよくするためのチェックビットのつけかた

例：Code (6, 3) のばあい

$$x_5=x_1\oplus x_3$$

$$x_6=x_2\oplus x_3$$

$$x_4=x_5\oplus x_6$$

とする．ここで⊕はケタ上りなしに加えることを意味する．すると S_3, S_2, S_1 の症候群が表わす符号語の2進数が誤り符号語の“番目”を指示する．

このほか，パリティチェックを2重にする法も利用される．

$L=16$ で訂正可能であるためには Code(21, 16) で，チェックビットは5コとなる．

いっぽうこの16ビットを4×4のマス目にならべ，各行，各列で偶数パリティになるよう1行目から $x_{17}\cdots\cdots x_{20}$，また1列目から $x_{21}\cdots\cdots x_{24}$ の8コのチェックビットを付加すると，さきの5ビットより3ビット多くなるだけで，情報の16ビットを2重にチェックすることができる．

ここでは，通信を誤りなく送り届けるための考え方をみてきた．誤りなく；つまり“正確さ”という概念は，情報の“質”的な側面を代表しているように思われるが，文字列を2元符号語に変換して送るばあい，チェックビットを付加するという“量”的な支援によって“正確さ”ないし“質”が保持ないし確保されるということは注目に価する．

6. 符号当りの情報量

文字の符号化にあたっては，文字の発生確率を考慮して，確率の大きい文字ほど長さの短い符号語（**最短符号語**）を割りあてる **Huffman の手法**は最短符号語の条件をみたすものである．

いま，M 種の文字

$$a_1, a_2, \cdots\cdots, a_M$$

を含む文字列で，各文字の発生確率が

$$p_1, p_2, \cdots\cdots, p_M$$

であるとき決められた最短符号語の長さを

（注） Huffman の符号化とは，前出の「いいあて法」で $\bar{q}$ (min) をうるための分割形式で紹介した第3次接近と同じ構造をもつ手法である．

$$l_1, l_2, \cdots\cdots, l_M$$

とすると，

1文字当りの平均符号長 l は

$$l=\Sigma p_i l_i \quad \text{bit}$$

であり，また，

1文字当りの平均情報量 H は

$$H=\Sigma p_i \lg 1/p_i \quad \text{bit}$$

となる．これより1符号語がうけもつ平均情報量は H/l となる．

ここで分子 H は確率が与えれば決定する関数，分母は，確率が与えられていても符号語の長さによって異なる関数である．

（i） いま $l_i(i=1, 2, \cdots\cdots, M)$ に最短符号語が割当てられたとすると $(H/l)=1$ であるから，それより長い符号語が割当てられるなら，一般に

$$H/l \equiv \beta(<1)$$

とするための l_i の分布は

H/l を最大化する条件—— Lagrange の未定係数法を適用して解くことができて

$$l_i=\frac{1}{\beta}\lg\frac{1}{p_i}$$

がえられる．

（ii） 他方 l_i の分布が与えられたばあいも H/l を最大化する条件は同じであるから

上式から p_i を求めると

$$p_i=2^{-\beta l_i}$$

がえられる．しかし，今度は β がまだ決まっていないから，これだけでは p_i の分布はきまらないが p_i に関する性質

$$\Sigma p_i=1$$

を適用すると

$$\Sigma 2^{-\beta l_i}=1$$

となる.

そこで β のかわりに

$$2^{\beta}\equiv W$$

を未知数とした

$$\Sigma W^{-l_i}=1$$

の高次方程式にもちこむと，この形式の高次方程式は，“正根がただ1コだけ存在する”ことが証明されている (Frobenius).

なお，5次以上の代数方程式は一般解が求められず，数値解（近似的にはグラフ解）で求めるしかない.

いずれにせよ，W が求められるから，これより

$$\beta=\lg W=\frac{\log W}{\log 2}$$

また p_i の分布は

$$p_i=2^{-l_i \lg W}=10^{-l_i \log W}$$

より求められる.

計算例

前出 $M=7$ のばあいの最短符号語の長さ $l_1=2, l_2=2, l_3=3, l_4=3, l_5=3, l_6=4, l_7=4$ を使って p_i を求めてみる.

方程式は

$$2W^{-2}+3W^{-3}+2W^{-4}=1$$

これは容易に解けて

$$W=2.00$$

これより

$$\beta=\lg W=1.00$$

こうして H/l の最大値は 1.00 となる．

これは p_i が与えられたとして求めた最短符号語を使った $H/l=0.99^*$ よりわずかだが“大きく”なる．

$$^*H=(1/28)(1/\log^2)\left[7\log\frac{28}{7}+6\log\frac{28}{6}+\right.$$

$$5\log\frac{28}{5}+4\log\frac{28}{4}+3\log\frac{28}{3}+2\log\frac{28}{2}$$

$$\left.+1\log\frac{28}{1}\right]=2.61$$

$$l=\frac{7}{28}\times2+\frac{6}{28}\times2+\frac{5}{28}\times3+\frac{4}{28}\times3+\frac{3}{28}\times3$$

$$+\frac{2}{28}\times4+\frac{1}{28}\times4=\frac{74}{28}=2.64$$

$$\therefore\quad H/l\doteqdot\frac{2.61}{2.64}\doteqdot0.99$$

この最大値がえられるときの p_i を前出の与えられた p_i を区別するために $\hat{p}_i$ を使うと

$$\hat{p}_1=\hat{p}_2=2^{-2}=0.2500$$

$$\hat{p}_3=\hat{p}_4=\hat{p}_5=2^{-3}=0.125$$

$$\hat{p}_6=\hat{p}_7=2^{-4}=0.0625$$

がえられる．

さて，(H/l) max になるばあいは，

p_i と l_i の間には，

$$\lg p_i=-\beta l_i$$

の関係があるから，ヨコ軸に l，タテ軸に p の対数をとった，片対数グラフでは (l_i, p_i) は（0，1）を通る右下りの傾斜 β の直線上にのる．

（A） 情報伝送速度

文字は符号語に変換して送るから，文字の送る速度は，各文字に割当て

られた符号語の長さと符号語1コ送る速さできまる．

しかし重要なのは，文字の送る速度，つまり単位時間当り送れる文字数でなく，単位時間当り送れる“情報量”であるとして，これを**情報（量）伝達速度**という．

情報源からの情報の発生や情報を記録するばあいについても，それぞれ**情報発生速度**および**情報記録速度**が同じように定義できる．

いま，送らねばならない文字の字種を M とし，各文字の発生確率を $p_1, p_2, \cdots\cdots, p_M$ とすれば，文字1コ当りの平均情報量 H は

$$H=\Sigma\ p_i \lg 1/p_i \qquad (\text{bit}/\text{コ})$$

各文字の伝道時間を $t_1, t_2, \cdots\cdots, t_M$ とすれば，文字1コ当りの平均伝送時間 t は

$$t=\Sigma\ p_i t_i \qquad (s/\text{コ})$$

これより情報伝送速度は

$$V\equiv H/t \qquad (\text{bit}/\text{コ})$$

となる．

さて，各文字を符号語に変換したときの長さを $l_1, l_2, \cdots\cdots, l_M$ (bit/コ) とすると，各文字の伝送時間 t_i は符号語の長さ l_i に比例するとしてよく

$$t_i=\tau l_i$$

ここで τ は単位長さ（1 bit）の符号語を送るに要する時間で，送る手段によってきまる定数（s/bit）である．

これを使うと情報伝送速度は

$$V=H/\tau\Sigma\ p_i l_i$$

ここで　$1/\tau\equiv v$ (bit/s)

とおけば，これは単位時間当り送れる符号語の bit 数となる．**bit 通信速度**とよぶ．

さらに　$\Sigma\ p_i l_i=l$ (bit)

と１文字当りの平均符号語の長さを使うと

情報伝送速度は

$$V=v(H/l)$$

となる.

伝送速度 V を大きくするくふう

これには２つの考え方がある.

（ｉ） bit 通信速度 v を大きくする。

これは送る手段によって決まる.

例えば，テレタイプでは 50，ファクシミリで１万，テレビでは 1000 万 (bit/s) など．このように v は勝手に変えるわけにはいかない.

（ii） H/l を大きくする.

これは最短符号語を採用すればよい.

一般に $(H/l)_{\max}=\beta\ (\leq 1)$ を採用したときの伝送速度 $V_{\mathrm{mxa}}=v\cdot\beta$ を**通信路容量**という.

なお，通信路容量がかえられる条件から求められた (l_i, p_i) のセットを**最速符号語**とよぶ.

計算例

（１） 字種 $M=2$，符号語の長さ $l_1=1, l_2=2$ としたときの最速符号語となる p_1 および p_2 を求めてみる.

まず方程式 $$W^{-1}+W^{-2}=1$$

を解いて $$W=(1\pm\sqrt{5})/2 \quad \fallingdotseq 1.618$$

これより $$\beta=\log W/\log 2=0.69$$

従って $$p_1=10^{-1\log W}=0.62$$

$$p_2=10^{-2\log W}=0.38$$

また $$V\max=0.69v$$

（2）　字種 M が2より大きく 10 より小さいばあいで l_i が1から 10 までの異なる長さが割当てられるばあいについて，W および p_i を求めるには算出表が利用できる．国沢清典『OR のための情報理論入門』日科技連　1978

例：$M=10$ で，$l_1=1, l_2=2, l_3=3, l_4=4, \cdots\cdots l_{10}=10$ のばあい，この算出表を使うと $W=1.9990, \beta \fallingdotseq 0.99928 \fallingdotseq 1.0$ と求められる．

（B）　伝送速度の概念拡張

情報伝送速度を最大化するために H/l を最大化することで解決した．このなかみは

分子：H で1文字がもつ平均情報量，

分母：l で1文字がもつ平均符号長，

であった．見方を変えれば，

情報量は文字のもつ“情報的”属性量であり，

符号長は文字を伝送する際の“コスト的”属性量である．

符号は長くなるほど，送るのに時間がかかる，つまり“時間的コスト”が余計にかかる．また符号語をレタリングではりつけて郵送するばあいでも1符号語（0や1）の単価が同じならやはり符号が長くなるほど“金銭的コスト”が余計にかかる．

つまり最大化の条件とは，コスト当りの情報量を最大にする問題を解いたことになる．

文字列のばあい

文字列で，1字づつ質問回答法によって知っていくばあい，1回の質問に答えてもらうのに料金 c 円支払うとすると，ある文字 a_i を知るための質問回数を q_i として，a_i を知るに要するコスト y_i は

$$y_i = cq_i \text{ 円}$$

文字列のすべての文字が入手し終った時点で，単位料金当り入手した情報量は，この文字列がもたらした全情報量を，支払総額で割ったもので，これは平均情報量 H を平均料金 $y=\Sigma p_i y_i=c\Sigma p_i q_i$ で割った値に一致する．つまり，料金当りの情報量は

$$\frac{H}{y}=\frac{H}{c\Sigma p_i q_i} \qquad \text{bit/円}$$

いま，全文をできるだけ安上りに入手するために Huffman の手法に従って質問するなら質問回数 q_i は符号語の長さ l_i に数値的に一致し，料金当り入手する情報量は最大になる．

同種商品のばあい

文字列のように有限の字種を含む文字がつぎつぎに発生する，または送られてくる時系列的な確率現象に対して，空間的な確率的現象にもこの考え方が使える．

いま，文字を商品におきかえる．すると字種は，同じ機能をもつ商品の“異なる銘柄”に対応させられる．すると各銘柄の商品が出まわっている程度を，各商品の存在確率で表わすと，この状況は空間的な確率現象とみられる．

一般の商品は大部分が同種で異なる銘柄という形式になっていることは，酒，タバコ，香水，などでふつうにみられる．これらは，普及品から高級品という区別で理解されている．

さて，同種の商品で異なる銘柄 M 種あるとき，それらの商品の出まわりかたを存在確率として $p_1, p_2, \cdots\cdots, p_M$，それらの商品の価格を $y_1, y_2, \cdots\cdots, y_M$ とすると，これらの全商品の市場における存在を情報量的にとらえると，全情報量に対する総価格の比，つまり価格当りの情報量を求めて

$$\frac{H}{y}=\frac{H}{\Sigma p_i y_i} \qquad \frac{\text{bit}}{\text{円}}$$

をこの商品の評価する際の指標にできる.

レタリングのばあいは1文字に対してある長さ l_i の符号語を割当てるので，そのコスト cl_i がここでのコスト y_i になる.

さて，一般に，商品の価格の決定には，その品質，デザイン，使い勝手，耐久性，……など，いろんな要素が反映するであろうが，ここでは，商品もその基本的な評価の観点として，出にくさ，出まわりにくさ，手にはいりにくさ，誰もがもっていない程度……などつまりは稀少性の指標とみられる“情報量”を重視するのも商品に対する態度の1つの基本となろう.

それは，人の購買行動は，個人個人としては，好みや要求のままに銘柄を選び価格も安いもの，高いものを購入しているのだが，全体としての売行きの結果をみると，全情報量に対して総価格を最小にするように行動しているという仮説が，多くのばあい基本的に成立することが知られているからである.

それは，「価格当り情報量を最大化する」という仮説である．もっとも，これを人々の選択する行動に関する仮説としなくても，人間が行動を選択するばあいこの仮説を1つの“基準”とみなし，これにもとづいた試算結果を，実際の購買行動と比較する形で理解し，解釈するという使い方にも役立つ.

いま，ある同一商品群について $H/y\equiv\alpha$ を求めておくいっぽう，H/l が最大になる条件を求めて $\lg 1/p_i=\beta y_i$ のとき $(H/l)\max=\beta$ であることから，価格当りの最大情報量が求められる．つまり，

(i)　p_i を既知（与件）とすれば，それに見合う価格 y_i の審査が可能になるし

(ii)　y_i を既知（与件）とすれば，それに見合う，望ましい存在確率 p_i が算出される.

なおここで $\beta=\alpha$ とおいて解けば，両者とも，全商品を考えたときの H

および y となり，これらは同じで“変更がない”のであるから，これにもとずいて

（i） 販売者側では，p_i の実績から，それに見合う新価格 y_i をつけるのが客へのサービスであり

（ii） 購入者側では，p_i の実態を調査して，現行の価格体系 y_i を批判するのが消費者活動の第1歩となる．

さらに，この考え方を拡張すれば，商品の種類が異なっても，それぞれの種類の商品が，それぞれ各種の銘柄をもち異なる出まわりかたをしているなら，単位価格当りの情報量の逆数をとって，それぞれの商品群について，bit 当り何円かという，情報量当りの換算価格が求められる．これによって商品をすべて情報量で共通化して，商品群の価格比較をする際にも利用できる．

これより割高の商品群とか割安の商品群などと換算価格として明示でき，商品群のランクづけができる．そして，そのような価格差が生ずることの原因や理由を考察するところまで踏みこめることになる．

計算例

（1） 本の価格

ある本を限定豪華本として 2000 部刷る．これの普及版を 10 倍の2万部刷る．

すると市場に出まわる確率は

豪華本について　$p_h=2000/22,000=2/22$

普及本について　$p_l=20,000/22,000=20/22$

これよりそれぞれの価格 y_h, y_l の比は

$$\frac{y_h}{y_l}=\frac{\log 1/p_h}{\log 1/p_l}=\frac{1.04}{0.04}\fallingdotseq 25.16$$

そこで $y_l=1,000$ 円とつければ $y_h\fallingdotseq 2.5$ 万円の定価が適正であると求め

られる.

（2）　レタリングの適正価格

あるメーカのある大きさ（サイズ）のレタリングをとりあげる.

銘柄はアルファベットで

$$A, B, C, D, \cdots\cdots, Z$$

また，使用確率は文章での発生確率をそのまま流用して，

$$p_A, p_B, \cdots\cdots, p_Z$$

このとき

価格は，字種によらず同一価格でありこれを y_0（円）とすると，現状における価格当りの情報量 α は

$$\alpha \equiv \frac{H}{y} = \frac{H}{\Sigma\, p_i y_i} = \frac{H}{y_0}$$

となる.

いっぽう $\beta \equiv (H/y)_{\max}$ は，モールス符号語の長さについて求められた β をそのまま採用すると $\beta \fallingdotseq 0.58$ であるから，これより $y_i = (1/\beta)\lg 1/p_i$ から基準価格が決まるが，ここで売上総額を同じにするために $\beta = \alpha$ として，適正な価格体系は

$$y_i = (1/\alpha)\lg 1/p_i$$

を採用して決めるのが合理的だということになる.

いま $M=26$ のばあい $H_1=4.12$ として，レタリング1枚 $y_0=50$ 円とすると

$$y_i = \frac{50}{4.12} \cdot \frac{\log 1/p_i}{\log 2}$$

そこで例えば

発生確率が最大の文字 e について $p_e=0.11$，最小の文字 z について $p_z=0.0018$ を使ってみると，文字 e の価格は 3.50 円に，また z 価格は

110.65円にするのが適正であるということになる．

50円を基準にすれば，よく使われる文字 e は，現行の価格の約1/14にすべきだし使われかたの少ない文字 z は2.2倍の高い価格にしてよいと提案できる．

（3） レポートの評価

教育の現場が行われるテストやレポートの目的としては，

① 現解力，論理力，推理力などを調べるタイプのもの（自然科学系）

② 情報の要約力，解釈力などを調べるタイプのもの（人文・社会系）

のほかにも

③ 手法・技法の適用（応用）能力を調べるタイプのもの

などがある．自然科学系の問題でも，期待する解答としては，

ⓐ 正解が1種しかないもののほか

ⓑ 正解が数種あるばあい，さらには

ⓒ 解の目標や条件をコトバで設定して，発想やアイディアを要求するもの，まで考えられる．

発想やアイディアとなると，どれだけ異なる種類の解答がとび出してくるか全くわからないし，またそれらについてどう評価するかは容易ではない．一般にはその道の専門家の判断にまかせることになるが，専門に分化されていない領域のテーマのばあいどう扱うかは難問である．

こんなばあい，解答を内容によって分類して，各分類に属する解答の“少ない”ほうを高く評価するという観点が設定できる．つまり情報量最大化の基準を採用するわけである．

これの事例として，文章をタテ書きするばあいの4種，ヨコ書きの4種について，これらを区別する適切な「ネーミング」を学生から募集してみたことがある．

提出枚数238に寄せられた代表的な呼称を，つぎの4種類に区分した．

①　「右行下降式」で代表されるもの（右向き下降，右走降下，東進南移，右読み下，……などを含む）32 枚

②　「右下式」(ライトダウン，ES，RD 東南，左右上下……などを含む）66 枚

③　「“横”右式」(左上横，左横下……などを含む）50 枚，

④　「その他」これには，英文式，英語式など，「語学・文章」系の命名や，プッシュホン式，タイプライター式など「機械」系の命名や，3時30 分式とか十一式など「数字・記号」系ほかなどが混在していて 50 枚となった.

これら 4 種類に区分されたときの所属枚数を使って求めた $-\log p$ をそのまま評価値としてみると①が 87点，② 56点，③ 42 点，④ 68 点となる．④の取扱いには問題が残るとしても，点数をそのまま順位づけて①をを A，②を C，③を D，④を B とおいてみることはできる．岡山『テストにおける出題のタイプと成積評価の検討』日本科学教育学会　1982

論文を評価するのに，論文を読んだとき編集者の“驚き”の程度を表わす数値の自然対数をとっは指標をとり入れた評価値の決定法が Physics Today 1967 年 7 月号に発表されている．ここで驚きの程度と評価の指標との関係の 1 例をあげると

−7：30 年前にわかっていたことだ

0：だからどうだというのだ

1：説明が足りんね

2：この議論にはついていけん

3：疑わしいね

4：本当かね

6：他の雑誌に投稿したほうが

7：天才かな，それとも.

といったぐあいである.

これは，科学上の業績を評価する目安として，Gillman が最小限の驚きを表わす基準値を導入しようとした先駆的研究を発展させたものであると断っている.

（C） 予言の情報量

事象が２種の a と b だけに分類され，このいずれかの事象が繰返し十分長く発生するとする.

これを記録すれば a と b からなる２元記号列となる. 以下，この記号列で考える.

いま，記号列の長さを N，このうち a が発生したコ数を N_a，b が発生したコ数を N_b とすると，a が発生した確率 $p=N_a/N$，b が発生した確率 $q=N_b/N$ $(=1-p)$

ここで $N=N_a+N_b$

記号列がもたらすエントロピー H は

$$H=p\lg\frac{1}{p}=q\lg\frac{1}{q}$$

いま，これらの記号発生に関して，すべて“予言”が行われたとする.

① a と予言した A 回のうち，あたったばあいを n_a，はずれたばあいを n_a' とすると的中率 $\alpha=n_a/A$，$(A=n_a+n_a')$ で予言 a がもたらすエントロピー E_a は

$$E_a=\alpha\lg\frac{\alpha}{p}+(1-\alpha)\lg\frac{1-\alpha}{q}$$

② b と予言した B 回のうち，あたったばあいを n_b，はずれたばあいを n_b' とすると的中率 $\beta=n_b/B$ $(B=n_b+n_b')$ で予言 b がもたらすエントロピー E_b は

$$E_b=\beta\lg\frac{\beta}{q}+(1+\beta)\lg\frac{1-\beta}{p}$$

これより全予言のエントロピー E は

$$E=\frac{A}{N}E_a+\frac{B}{N}E_b$$

ここで予言がすべて完全なら,

① a の予言について的中率 $\alpha=1$ だから　$E_a=\lg 1/p$,

② b の予言についても的中率 $\beta=1$ だから $E_b=\lg 1/q$

これより全予言のエントロピー E_1 は

$A=N_a, B=N_b$　だから

$A/N=p, B/N=q$　となり

$$E_1=p\lg\frac{1}{p}+q\lg\frac{1}{q}$$

つまり事象がもたらすエントロピーに一致する. 斉藤嘉博『予測』日科技連　1970.

アンケート調査への応用

いま N 人にアンケート調査を行う. 回答は yes, no の2種に限るとする.

アンケートを1人づつ行って, その回答を記録すれば2元記号列となる.

これがもたらすエントロピー E_1 は

$$E_1=p\lg\frac{1}{p}+q\lg\frac{1}{q}$$

ここで　　$p=N_a/N, q=N_b/N(=1-p)$

N_a は a(yes) の, N_b は b(no) のコ数である.

さて, 回答者を予言者でもあるとみると, ある質問に対して a と回答したとすると, それは a と予言して回答 a を発生させたと解釈できるの

で，a の予言について的中率は1のばあいとして，その予言・回答者がもたらすエントロピー $E_a=\lg 1/p$，また b と回答したばあいはその予言・回答者がもたらすエントロピー $E_b=\lg 1/q$ となる．

いま i 番目の質問について $p=q$ であれば，a の予言・回答者も b の予言・回答者もそのもたらす情報量は 1 bit で同じである．ただし，その回答の内容は yes と no でその意味はへだたっている．つまり情報量は同じでも回答の意味内容からみれば，はっきり相違する．これによって a 予言・回答者と b 予言・回答者との2つのグループにわけられる．

また，別の j 番目の質問で，再び $p=q$ であれば，この質問の観点でも2つのグループにわけられる．

そしてこれらの分類はいずれも，情報量的にみて“公平”なわけかたになっている．

さて，k 番目の予言・回答者が，質問 i で a または b のいずれか，別の質問 j でも a または b のいずれかの予言・回答するから，その組合わせとして，予者・回答のパタンは aa, ab, ba, bb の4種になる．

一般に多くの m コの質問について，すべて $p=q$ がえられるなら，情報量的にみて等しい条件を満たしながら，予言・回答が異なるパタンによって最大 2^m 種にグループわけできる．

N 人の集団を，何ら分類せず，そのままにしておけば，その集合がもたらす情報量は0であるが，これを，ある観点で K 種に分類し，i 番目の分類に属する人数を N_i とすると，

この分類によってエントロピー

$$H=\sum_{i=1}^{K} p_i \lg 1/p_i \qquad (p_i=N_i/N)$$

がもたらされる．

従って分類することは集団のエントロピーを大きくすることを意味する．

従って分類では，種数 K が同じであってもエントロピーが大きくなるような観点での分類がそれだけ有効であり，意味があるといえる．

また，エントロピーは，その分類の“よさ”を示す指標となるいっぽうで，よい分類が保証されたとき，分類のなかみもそれに伴って注目に値するものとなる．

つまり単なる集団としての特性をエントロピーで吟味するだけにとどまらず，集団を構成する各個体について，それぞれがどの特性をもつグループに所属するかを知るのに役立てることができる．

これの具体例は，あとで事例3で示す．

また i 番目の質問に予言・回答 a をしたとすると，もたらす情報量は $\lg 1/p_i$，また予言・回答 b をしたら，情報量 $\lg 1/q_i$ をもたらしたことになる．

すると質問 $i(i=1, 2, \cdots\cdots, m)$ に対してすべて予言・回答したとき，もたらした全情報量は，これら情報量の和

$$I=\sum_j \lg 1/p_i+\sum_i \lg 1/q_j$$

ここで $\sum_i$ は a と予言・回答した質問番号 i のすべて，$\sum_j$ は b 予言・回答した質問番号 j のすべてについて加算することを意味する．

このように情報量は，予言・回答という情報の指標（以下得点とよぶ）とみられるが，この得点は，各質問に対して，予言・回答数が相対的に少ないグループに属するほど“大きく”なる．つまり多くの予言・回答者と違う予言・回答をするほど得点は高くなる．そして，このような傾向が著しく表われるほど，その予言・回答者は，その持前の“未来分析の感覚”において精神的に“若い”とみることができる．

「若さ」をどう定義するかについては，あらためて考察し，議論を重ねなければならないのだが，ここでは，基本的な若さの特質として“多数や現状に対する拒否願望ないし態度”を全面にかかげておくことにする．

そこで全情報，つまり総得点が高いほど“若い”ということになる．

これの具体例を，このあと事例4で示す．

（事例 3）　予言・回答のグループわけ

朝日新聞社が「1970年代の日本」を予測するアンケート調査を行った．そのデータが1970年1月1日の紙上に発表された．

質問は「……ですか」とか「……になるか」という種類の20コで，これにyes, noで回答する形式である．

調査された対象は，70年代のにない手として選ばれた100人である．

データは，これらの人たちの回答をyesなら○印，noなら×印で20コからなる○×の記号列で表わしてある．

表 15　20の質問に対する回答数および情報量計算

質問コード i	回答 ○ a_i	回答 × b_i
1	32	68
2	69	31
3	27	73
4	33	67
5	09	91
6	27	73
7	34	66
8	45	55
9	09	91
10	12	88
11	09	91
12	22	78
13	71	29
14	31	69
15	26	74
16	32	68
17	18	82
18	56	44
19	25	75
20	64	36

$\Sigma a_i=651$
$\Sigma b_i=1349$
$N=2000$

質問コード i	情報量 ○ $\lg 1/p_i$	情報量 × $\lg 1/p_i$	平均情報量
1	1.644	0.556	0.904
2	0.535	1.690	0.893
3	1.889	0.454	0.841
4	1.600	0.578	0.915
5	3.474	0.136	0.436
6	1.889	0.454	0.841
7	1.556	0.599	0.925
8	1.152	0.862	0.993
9	3.474	0.136	0.436
10	3.059	0.184	0.529
11	3.474	0.136	0.436
12	2.184	0.358	0.760
13	0.494	1.786	0.869
14	1.690	0.535	0.893
15	1.943	0.434	0.827
16	1.644	0.556	0.904
17	2.474	0.286	0.680
18	0.837	1.184	0.990
19	2.000	0.415	0.811
20	0.644	1.474	0.943

なお，回答のなかには○とも×とも回答せず？（保留）が16コ含まれているが，1つの質問について最大3コにとどまる程度に分散しているので，それぞれの回答の○と×のうち回答数の少ないほうに入れて加算した．

このようにデータを多少修整したあと，質問1から 20 まで順序に，その○，×の回答数は，表 15 に示す.

ここで　$N_a=\Sigma a_i=651$,

$N_b=\Sigma b_i=1349$, $N=2000$

である.

さてこのデータ (a_i, b_i) $(i=1, 2, \cdots\cdots 20)$ をみると $a_i=b_i$ というのはみつからないが，質問8に対する回答数が最も接近している.

この質問は「自衛隊が出動する騒乱が起るか」で，○が 45 各（？の2名を含む)，×が 55 名である．つぎに接近しているのは，質問 18 で「恋人が街頭で Kiss するようになるか」で，○が 56 名，×が 44 名である.

なお，質問8は内容的には“硬派”の現象であり，これに対し質問 18 は“軟派”の現象であるという対比，また，回答数も，前者では×が○より多少多く，後者では○が×より多少多いが，回答数比は，前者と後者で逆転するという対比を示すことなど偶然とはいえない気がするほどである. それはさておき，エントロピーの大きいこれら2種の回答群を取り出し質問 8，質問 18 に対する個人の回答を，この順にならべて

yes　yes ⟶ ○○……T_1

yes　no ⟶ ○×……T_2

no　yes ⟶ ×○……T_3

no　no ⟶ ××……T_4

と4つのタイプにわける．そして，これら4種は，該当する質問の内容とその回答に，回答者の予測傾向を反映しているので，上から革新派，保守派，楽観派そして悲観派とよぶ.

ここで，回答群のもつエントロピーをみておこう（表 15).

質問8は 0.99，質問 18 も 0.99 でほぼ同値.

エントロピーの最小なのは質問 5，9，および 11 でいずれも 0.44 で，

大きい値の半分以下におちこんでいる.

なお，このエントロピーをもつ3種の質問とは：――

質問5は「部課長は部下が選ぶようになるか」であり，質問9は「大規模な戦争が起るか」であり，質問 11 は「マイカー族は減るか」である．これらは，いずれも yes 回答が，全回答の1割弱にとどまっているというわけである.

さて4種のタイプ T_1, T_2, T_3, T_4 に属する人々のリストを示す（表 16).
岡山『人間の分類と評価のテクニック』蟻塔　1975―2.

（事例 4） 予言・回答の「若さ」指数

再び前出の朝日新聞社のアンケート調査（'70―1―1）を利用する.

各質問 $i(i=1, 2, \cdots\cdots, 20)$ についての○，×の回答数 a_i と b_i のデータから，

$$p_i=a_i/N,\quad q_i=b_i/N$$

を求め，さらに $\lg 1/p_i$ および $\lg 1/q_i$ を求める（表 15).

これを使って，予言・回答者の 20 コの回答パタンと照合しながら上の表から情報量を求めて総和をとる．こうして求めた全情報量を大きさの順にならべる（表 17).

なお，20 コの予言・回答で，情報量の大きいほうばかりが選択されたとすると全情報量は 41.25（最大値）となる．さすがこの値になるような予言・回答者はいなかった．また，逆に，情報量の小さいほうばかりを拾ってくると 9.21（最小値）がえられるが，この値を示す予言・回答者は110位に存在する.

また，これら最大値や最小値を使って全情報量を基準化して多少イメージしやすい指標を導くことも可能であるが今回は省略する.

表 16　70 年代をになう人，100 名の回答パターン

T_1			T_2		
飛鳥田一雄	54	横浜市長	天谷　直弘	44	通産省企画室長
渥美　和彦	41	東大教授（医用電子学）	鮎沢　啓夫	42	九大教授(微生物学)
五木　寛之	37	作家	池田　芳蔵	56	三井物産常務
井上　　修	49	日米水産社長（山口県）	石原　　俊	57	日産自動車専務
井上　光晴	43	作家	いずみたく	39	作曲家
上野　　孝	28	動力車労組青年部長	犬丸　　直	46	文部省人事課長
牛尾　治朗	38	ウシオ電機社長	今堀　和友	49	東大教授（生物物理学）
内田　　勝	34	少年マガジン編集長	岩沢　正二	56	住友銀行常務
梅垣　哲郎	50	電通常務	宇佐美忠信	44	全繊同盟書記長
浦山　桐郎	39	映画督監	江藤　　淳	36	文芸評論家
大田　昌秀	44	琉球大教授(社会学)	大平　正芳	59	自民党
金井美恵子	22	詩人	尾崎和三郎	55	松下電器報道部長
木村三千雄	44	国際農村青年協会長(新潟県)	角本　良平	49	交通評論家
草柳　大蔵	45	ルポライター	上坂　冬子	39	評論家
コシノジュンコ	30	デザイナー	高坂　正尭	35	京大助教授（国際政治学）
三枝　守雄	56	石川島播磨重工取締役	河野　洋平	32	自民党
鈴木　義雄	44	スズヤ社長	香山　健一	36	学習院大助教授（社会工学）
関根　智明	43	慶大助教授（管理工学）	庄司　　薫	32	作家
高橋　和巳	38	作家	菅谷　隆介	53	日本興業銀行取締役
田淵　節也	46	野村証券常務	田川　誠一	51	自民党
堤　　清二	42	西武百貨店社長	田中　角栄	51	自民党
勅使河原霞	37	生け花家	中内　　功	47	ダイエー社長
戸田　修三	46	中大教授（商法）	中曽根康弘	51	自民党
永井　道雄	46	東京工大教授（教育学）	南雲　仁一	43	東大教授(情報工学)
橋本　道夫	45	厚生省公害課長	服部　盛栄	59	三菱商事常務
藤本　陽一	44	早大教授（原子物理学）	藤吉　次英	56	東レ副社長
松本三之介	43	東京教育大教授（日本史）	不破　哲三	39	共産党
宮本　憲一	39	大阪市大助教授（財政学）	松下　圭一	40	法政大教授(政治学)
森　　一久	43	原子力産業会議事務局長	矢野　絢也	37	公明党
吉田　直哉	38	NHK ディレクター	渡部　正郎	50	警察庁審議官
和田　　誠	33	イラストレーター			
31 名			30 名		

によるグループ別リスト（全）

T_3			
浅利	慶太	36	演出家
池田	大作	41	創価学会会長
石原	信吾	51	虎の門病院事務部長
石山	四郎	50	経済評論家
今井	大宗	55	八幡製鉄常務
岩間	和夫	50	ソニー専務
内田	忠夫	46	東大教授（計量経済学）
岡崎	令治	39	名大教授（生化学）
加藤	秀俊	39	京大助教授(社会学)
河上	民雄	41	社会党
木島	昴	41	開業医
近衛	忠煇	30	日赤外事部主事
坂本	義和	42	東大教授（国際政治学）
佐藤	八次	48	東京都広報室長
下河辺	淳	46	経済企庁調査官
外村	隆	50	法務総合研究所研修第一部長
鍋島	直紹	57	自民党
橋本	恕	43	外務省中国課長
槙枝	元文	48	日教組書記長
三島由紀夫		44	作家
宮崎	勇	46	経企庁調査官
森	秀太郎	55	トヨタ自工取締役
安村	美博	41	千葉大助教授（微生物学）
吉永小百合		24	女優
渡辺	美佐	41	渡辺プロ副社長
25 名			

T_4			
石橋	正嗣	45	社会党
市村	真一	44	京大東南ア研究センター所長
岩田	隼	49	日立製作所計算制御部長
衛藤	瀋吉	46	東大教授（国際関係論）
小田	実	37	ベ平連
北小路	敏	33	社会運動家
黒川	紀章	35	建築家
小長井良浩		35	総評弁護団
佐々木良作		54	民社党
高木	文雄	50	大蔵省審議官
高畑	敬一	40	電機労連大阪地協議長
高宮	昇	53	東芝常務
中村	悌次	50	防衛庁第5幕僚室長
我妻	尭	39	愛育病院産婦人科部長
14 名			

五十音順・敬称略・数字は年齢（'70・1・1 現在）

表 17　70年代をになう人，100名の回答パターンによる全情報量の順位リスト（全）

氏名	年齢	職業	情報量
中内　功	47	ダイエー社長	29.71
吉田直哉	38	NHK ディレクター	22.73
五木寛之	37	作家	22.43
渥美和彦	41	東大教授（医用電子学）	21.87
宇佐美忠信	44	全繊同盟書記長	21.54
角本良平	49	交通評論家	21.23
上坂冬子	30	評論家	20.99
石山四郎	50	経済評論家	20.79
岩間和夫	50	ソニー専務	20.75
不破哲二	39	共産党	20.66
木島　昴	41	開業医	20.65
飛鳥田一雄	54	横浜市長	20.57
下河辺　淳	46	経企庁調査官	20.45
宮崎　勇	46	経企庁調査官	19.77
森　一久	43	原子力産業会議事務局長	19.74
永井道雄	46	東京工大教授（教育学）	19.70
関根智明	43	慶大助教授（管理工学）	19.31
高橋和巳	38	作家	19.03
草柳大蔵	45	ルポライター	18.91
牛尾治朗	38	ウシオ電機社長	18.75
高畑敬一	40	電機労連大阪地協議長	18.73
北小路　敏	33	社会運動家	18.69
外村　隆	50	法務総合研究所修第一部長	18,69
矢野絢也	37	公明党	18.63
三島由紀夫	44	作家	18.38
田中角栄	51	自民党	18.27
上野　孝	28	動力車労組青年部長	18.09
内田忠夫	46	東大教授（計量経済学）	17.79
香山建一	36	学習院大助教授（社会工学）	17.51
木村三千雄	44	国際農村青年協会長（新潟県）	17.40

氏名	年齢	職業	得点
井上　修	49	日本水産社長（山口県）	17.33
鈴木義雄	44	スズヤ社長	17.27
橋本道夫	45	厚生省公害課長	17.27
加藤秀俊	39	京大助教授（社会学）	17.17
尾崎和三郎	55	松下電器報道部長	16.86
藤本陽一	44	早大教授（原子物理学）	16.79
高坂昭尭	35	京大助教授（国際政治学）	16.50
池田大作	41	創価学会会長	16.46
高宮　昇	53	東芝常務	16.41
市村真一	44	京大東南ア研究センター所長	16.40
石橋政嗣	45	社会党	16.22
松下圭一	40	法政大教授（政治学）	16.21
和田　誠	33	イラストレーター	16.20
宮本憲一	39	大阪市大助教授（財政学）	16.19
浦山桐郎	39	映画監督	16.08
大平正芳	59	自民党	15.85
勅使河原　霞	37	生け花家	15.51
コシノジュンコ	30	デザイナー	15.50
金井美恵子	22	詩人	15.49
河野洋平	32	自民党	15.47
佐藤八次	48	東京都広報室長	15.47
渡辺美佐	41	渡辺プロ副社長	15.27
菅谷隆介	53	日本興業銀行取締役	15.17
森　秀太郎	55	トヨタ自工取締役	15.08
佐々木良作	54	民社党	14.88
今堀和友	49	東大教授（生物物理学）	14.77
浅利慶太	36	演出家	14.75
石原　俊	57	日産自動車専務	14.66
犬丸　直	46	文部省人事課長	14.65
池田芳蔵	58	三井物産常務	14.61
黒川紀章	35	建築家	14.47
小田　実	37	ベ平連	14.45
戸田修三	46	中大教授（商法）	14.22

氏名	年齢	職業	
梅垣哲郎	50	電通常務	14.07
江藤淳	36	文芸評論家	13.96
衛藤瀋吉	46	東大教授（国際関係論）	13.96
藤吉次英	56	東レ副社長	13.93
橋本恕	43	外務省中国課長	13.91
堤清二	42	西武百貨店社長	13.64
石原信吾	51	虎の門病院事務部長	13.57
三枝守雄	56	石川島幡磨重工取締役	13.46
いずみたく	39	作曲家	13.02
内田勝	34	少年マガジン編集長	12.97
庄司薫	32	作家	12.97
服部盛栄	59	三菱商事常務	12.97
鍋島直紹	57	自民党	12.89
鮎沢啓夫	42	九大教授（微生物学）	12.83
近衛忠煇	30	日赤外事部主事	12.82
渡部正郎	50	警察庁審議官	12.76
中曽根康弘	51	自民党	12.75
我妻尭	39	愛育病院産婦人科部長	12.74
槙枝元文	48	日教組書記長	12.69
河上民雄	44	社会党	12.63
坂本義和	42	東大教授（国際政治学）	12.46
南雲仁一	43	東大教授（情報工学）	12.02
小長井良浩	35	総評弁護団	12.01
田渕節也	46	野村証券常務	11.95
田川誠一	51	自民党	11.68
今井大宗	55	八幡製鉄常務	11.62
井上光晴	43	作家	11.48
松本三之介	43	東京教育大教授（日本史）	11.48
中村悌次	50	防衛庁第5幕僚室長	10.99
岩沢正二	56	住友銀行常務	10.97
大田昌秀	44	琉球大教授（社会学）	10.94
高木文雄	50	大蔵省審議官	10.92

岩田 隼	49	日立製作所計算制御部長	10.80
安村美博	41	千葉大助教授（微生物学）	10.72
天谷直弘	44	通産省企画室長	10.70
吉永小百合	24	女優	10.23
岡崎令治	39	名大教授（生化学）	9.21

敬称略　数字は年齢（'70・1・1 現在）